AUSTRALIA

Please Search into
Case #
10/640,647

Patent Office
Canberra

I, GAYE TURNER, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PP 0422 for a patent by THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH filed on 18 November 1997.

I further certify that the above application is now proceeding in the name of AMRAD OPERATIONS PTY LTD pursuant to the provisions of Section 113 of the Patents Act 1990.

WITNESS my hand this
Twentieth day of July 2001

G Turner.

GAYE TURNER
TEAM LEADER EXAMINATION
SUPPORT AND SALES

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

Regulation 3.2

Amrad Operations Pty Ltd

~~The Council of the Queensland Institute of Medical Research~~

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

for the invention entitled:

"Novel molecules"

The invention is described in the following statement:

NOVEL MOLECULES

FIELD OF THE INVENTION

The present invention related generally to novel molecules and more particularly novel proteinaceous molecules involved in or associated with regulation of cell activities and/or viability. The present invention is particularly directed to novel serine proteinases and a novel kinase and to derivatives, agonists and antagonists thereof. In one embodiment, the present invention provides a novel serine proteinase, referred to herein as "HELA2" or "testisin", which has a role in spermatogenesis, in suppressing testicular cancer and as a marker for non-testicular cancers.

Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide and amino acid sequences referred to in the specification are defined at the end of the subject specification.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

BACKGROUND OF THE INVENTION

The rapidly increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. This is particularly the case in the area of cell regulation leading to a greater understanding of the events leading to or involved in cancer, development of acquired immunodeficiency disease syndrome (AIDS), neurological disorders, heart disease, tissue graft rejection and infertility amongst many other conditions.

Two particularly important classes of molecules are the proteinases and kinases.

Proteinases play important roles in a number of physiological and pathological processes such

as proteolytic cascades involved in blood coagulation, fibrinolysis and complement activation as well as cleavage of growth factors, hormones and receptors, the release of bioactive molecules and processes involving cell proliferation and development, inflammation, tumour growth and metastasis. Of particular significance are the cellular proteinases, or those proteinases synthesized in cells and tissues which serve to activate or deactivate proteins responsible for performing specific functions. These proteinases may be found outside the cell, within the cell or may be present on the cell surface.

Serine proteinases are particularly important. These proteinases are characterised by a mechanism involving serine, histidine and aspartate amino acids in the serine proteinase active site. Members of the serine proteinase family which play important roles in a range of cellular functions and which have demonstrated causative roles in human diseases include tissue-type plasminogen activator and thrombin (thrombosis and blood clotting), urokinase-type plasminogen activator (cancer and metastasis), trypsin and elastase (emphysema and liver disease) and angiotensin converting enzyme (hypertension).

A serine proteinase is also implicated in TNFα degradation and soluble TNF-receptor (p75) release by THP1 cells (Vey *et al. Eur. J. Imm. 26,* 2404-2409, 1996). Serine proteinases have been implicated in the activation of macrophages (Nakabo *et al. J. Leukocyte Biol. 60,* 328-336, 1996), in nuclear lamin degradation in apoptosis (McConkey *et al. J.Biol. Chem., 271,* 22398-22406, 1996), in prostaglandin-E2 induced release of soluble TNF receptor shedding (Choi *et al. Cellular Immunology 170,* 178-184, 1996), in PAF sysnthesis (Bussolino *et al. Eur. J. Immunol. 24,* 3131-3139, 1994), and in the proteolysis of IkB, a regulatory molecule important in signal transduction and apoptosis. Release of serine proteinases known as granzymes is central to CTL killing and many of the substrates cleaved by granzymes are also cleaved by cellular proteinases (for example, IL-1β is a substrate for Granzyme B as well as the cysteine proteinase, interleukin 1β-converting enzyme (ICE)). Granzyme A, a serine proteinase with Arg-amidolytic activity, has been reported to induce the production of IL-6 and IL-8 in lung fibroblasts (Sower *et al. Cellular Immunology 171,* 159-163, 1996) and cleaves IL-1β to a 17kD mature form that is biologically active.

Kinases are a large group of molecules, many of which regulate the response of cells to external stimuli. These molecules regulate proliferation and differentiation in eukaryotic cells frequently *via* signal transduction pathways.

The identification of new serine proteinases and kinases permits the development of a range of derivatives, agonists and antagonists at the nucleic acid and protein levels which in turn have applications in the treatment and diagnosis of a range of conditions such as cancer, inflammation, neurological disorders amongst many other conditions including conditions which initiate or promote apoptosis such as viral infection, old and drug abuse. One particularly useful serine proteinase HELA2 (testisin) identified in accordance with the present invention is involved in spermatogenesis, testicular cancer and as a marker for non-testicular cancer.

SUMMARY OF THE INVENTION

One aspect of the present invention provides a novel molecule in isolated form involved in or associated with regulation of cell activity and/or viability.

Another aspect of the present invention contemplates a novel serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence, at least a portion of which, is capable of being amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ ACACTTAAGAXIGGICCICCIC/GT/AXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

Yet another aspect of the present invention is directed to an isolated serine proteinase comprising the amino acid sequence substantially set forth in SEQ ID NO:4 or an amino acid sequence having at least about 50% similarity to all or part thereof. This serine proteinase is referred to

herein as a short isoform (S) of "HELA2" or "testisin".

Still another aspect of the present invention relates to an isolated serine proteinase comprising the amino acid sequence substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least about 50% similarity to all or part thereof. This serine proteinase is referred to herein as a long isoform (L) of HELA2 (testisin).

Still yet another aspect of the present invention provides an isolated serine proteinase comprising an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity to all or part thereof. This serine proteinase is referred to herein as "ATC2".

Even yet another aspect of the present invention is directed to a serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or a nucleotide sequence having at least 50% similarity to all or part thereof or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions.

Another aspect of the present invention relates to a serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity to all or part thereof or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions.

Still another aspect of the present invention provides a serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity to all or part thereof or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions.

Another embodiment of the present invention is directed to a kinase in isolated form comprising

an amino acid sequence substantially as set forth in SEQ ID NO:10 or having 50% amino acid similarity to all or part thereof. This kinase is referred to herein as "BCON3".

In a related embodiment, the kinase comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity to all or part of the nucleotide sequence set forth in SEQ ID NO:9 or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions.

The present invention further provides an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:3 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:3 under low stringency conditions.

Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:5 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:5 under low stringency conditions.

Still another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:7 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:7 under low stringency conditions.

Even still another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:9 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:9 under low stringency conditions.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a representation showing (A) schematic and (B) hydrophobicity plot of the HELA2 amino acid sequence.

Figure 2 is a photograph representation showing Northern blot analysis of HELA2.

Total RNA was isolated from HeLa cells and HeLa cells stably transfected with a PA1-2 expressing construct (S1a cells). S1a cells had been treated with TNF and cycloheximide for the times shown. Ten μg of total RNA was electrophoresed on a denaturing agarose gel, transferred to Hybond N and fixed by UV irradiation. DNA for probing was recovered by EcoR1 digestion of the pGEM-T cloned RT-PCR product (450 bp). Radiolabelling was performed by the random priming method. Blots were washed to a stringency of 0.2 x SSC and 0.1% w/v SDS.

Figure 3 is a photographic representation showing Northern analysis of BCON3. The method is as described in the legend for Figure 2.

Figure 4 is a diagrammatic representation showing the amino acid sequence of HELA2 (testisin). The putative signal sequence, light chain, heavy chain and transmembrane domains are as indicated. The catalytic amino acids, His, Asp and Ser are as designated. Insertion of Tyr-Ser (YS) 4 amino acids after the catalytic His is found in the long isoform of testisin and is due to alternative mRNA splicing.

Figure 5 is a diagrammatic representation of plasmid constructs pBluescriptHELA2(S) and pBluescriptHELA2(L) containing full length cDNAs for testisin (short isoform (S)) and testisin (long isoform (L)), respectively.

Figure 6 is a diagrammatic representation of plasmid constructs pQET(20-295)N and pQET(20-295)C, wherein the hydrophobic residues of testisin were removed and the remaining sequences cloned into pQE prokaryotic expression plasmids.

Figure 7 is a photographic representation of: (A) Western blot analysis showing purification of recombinant HELA2 (testisin) from *E. coli.* The purified HELA2 (testisin) is indicated by the arrow in the eluate fractions. Some HELA2 (testisin) is also found in the wash fractions as the affinity matrix was not used in excess. His-N21 is one clone containing the amino-terminal His tag, and clones His-C21, His-C22 and His-C23 are three different clones with the carboxy-terminal His tag. (B) Western blot of native and denatured recombinant HELA2 (testisin) probed with Clontech anti-His tag-antibody. The 32kD band shown by the arrow is HELA2 (testisin). HELA2 (testisin) is not detected in the denatured samples as it appears that denaturation with urea destroys the His epitope recognised by the monoclonal antibody.

Figure 8 is a representation of the amino acid sequence of HELA2 (testisin) showing the regions of the molecule selected for generation of peptide antigens.

Figure 9 is a representation of an ELISA assay showing typical antibody response to peptide antigen. The peptide was coated onto the plate (25 ng) and following blocking rabbit serum was applied at dilutions of 1:10, 1:100, 1:1000 and 1:10000. Bound antibody was detected with peroxidase conjugated sheep anti-rabbit antibodies and visualised with substrate ABTS. Absorbance readings were at 415 nm.

Figure 10 is a diagrammatic representation of eukaryotic expression constructs, pcDNA3-Test(S-C), pcDNA3-Test(L-C) and pcDNA3-Test(1-297)L-C.

Figure 11 is a photographic representation showing Clontech Master RNA blot of the tissue distribution of HELA2 (testisin) in RNA from 50 different normal tissues. (A) probed with a HELA2 (testisin) specific probe. (B) probed with a control probe, BCON3, which is ubiquitously expressed. Control RNA is found in the bottom row.

Figure 12 is a photographic representation of agarose gel of PCR products generated by amplification of HELA2 (testisin) cDNA in prevasectomised and post-vasectomised ejaculate specimens. The HELA2 (testisin) PCR product is 464bp and the β2-macroglobulin product is 250 bp.

Figure 13 is a representation providing a summary of expression pattern of testisin mRNA as determined from Clontech Master blot, Northern blot of tumour cell lines and EST database matches. (+) indicates that HELA2 (testisin) is expressed; (-) indicates that HELA2 (testisin) is not expressed and (+/-) indicates that HELA2 (testisin) is barely detectable.

Figure 14 is a representation showing spread of normal metaphase chromosomes showing bright dots where HELA2 (testisin) is expressed at 16p13.3.

Figure 15 is a diagrammatic representation showing a map of cosmids which contain DNA sequences which hybridise to HELA2 (testisin) cDNA.

Figure 16 is a diagrammatic representation of the human genome map showing the location of HELA2 (testisin) between P-MWH2A and CY23HA.

Figure 17 is a photographic representation of northern blot analysis of HELA2 (testisin) mRNA showing signals in normal testis of 4 patients and absence of signal in the tumours of these patients.

Figure 18 is a diagrammatic representation showing a genomic map of HELA2 (testisin) compared with the closest serine proteinase homologue, prostasin. Predicted and experimental determined intron/exon boundaries are as shown. Exons are boxed and marked with a Roman numeral. Introns are lines and marked with a letter. ND indicates not yet determined.

Figure 19 is a representation of the DNA sequence of Intron C and flanking exons showing where alternative mRNA splicing occurs to generate the two isoforms of HELA2 (testisin).

Figure 20 is a representation showing location of PCR primers synthesised for mutational analysis of the HELA2 (testisin) gene by analysis of genomic DNA and SSCP analysis.

Figure 21 is a representation showing DNA and amino acid sequence of mouse HELA2 (testisin).

A summary of the SEQ ID NOs used throughout the specification is presented in Table 1.

SEQ ID NO	DESCRIPTION
1	PCR primer sequence
2	PCR primer sequence
3	Nucleotide sequence of short form of HELA2
4	Amino acid sequence of short form of HELA2
5	Nucleotide sequence of long form of HELA2
6	Amino acid sequence of long form of HELA2
7	Nucleotide acid sequence of ATC2
8	Amino acid sequence of ATC2
9	Nucleotide acid sequence of BCOM3
10	Amino acid sequence of BCOM3
11	Primers used to generate amino terminal tagged protein
12	Primers used to generate amino terminal tagged protein
13	Primers used to generated carboxy-linked terminal protein
14	Primers used to generated carboxy-linked terminal protein
15	Peptide antigen T20-33
16	Peptide antigen T46-63
17	Peptide antigen T175-190
18	Forward primer
19	Reverse primer
20	Forward primer
21	Reverse primer
22	Forward primer
23	Reverse primer
24	Serine proteinase activation motif

A list of single and three letter abbreviations for amino acid residues is presented in Table 2.

TABLE 2

Amino Acid	Three-letter Abbreviation	One-letter Symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Any residue	Xaa	X

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention is predicated in part on a genetic engineering approach to identify nucleotide sequences encoding serine proteinases or kinases. The genetic engineering approach is based on the use of degenerate primers corresponding to conserved regions of serine proteinases (amino acids flanking His- and Ser- residues) to amplify gene fragments spanning these regions for cDNA, using low stringency reverse transcriptase-polymerase chain reaction (RT-PCR).

This technique has been successfully used, in accordance with the present invention, to identify serine proteinases and kinases useful in modulating cell activity and viability including modulating spermatogenesis, acting as tumour suppressors and acting as a marker for non-testicular cancers.

Accordingly, one aspect of the present invention provides a novel molecule in isolated form involved in or associated with regulation of cell activity and/or viability.

More particularly, the present invention contemplates a novel serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence, at least a portion of which, is capable of being amplified by polymerase chain reaction (PCR) using the following primers:

5′ ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3′ [SEQ ID NO:1]; and

5′ ACACTTAAGAXIGGICCICCIC/GT/AXTCICC3′ [SEQ ID NO:2];

or a complementary form of said primers.

In a particularly preferred embodiment, the isolated serine proteinase comprises the amino acid sequence substantially set forth in SEQ ID NO:4 or an amino acid sequence having at least about 50% similarity to all or part thereof. This serine proteinase is referred to herein as a short isoform of "HELA2" or "HELA2 (testisin)". The terms "HELA2" and "testisin" are used

interchangedly throughout the subject specification to refer to the same molecule.

In another preferred embodiment, the amino acid sequence of the serine proteinase is substantially as set forth in SEQ ID NO:6 or an amino acid sequence having at least about 50% similarity to all or part thereof. This serine proteinase is the long isoform of HELA2 or HELA2 (testisin).

Yet another preferred embodiment of the present invention provides an amino acid sequence substantially as set forth in SEQ ID NO:8 or an amino acid sequence having at least about 50% similarity to all or part thereof. This serine proteinase is referred to herein as "ATC2".

Another aspect of the present invention relates to a serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:3 or a nucleotide sequence having at least 50% similarity to all or part thereof or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:3 under low stringency conditions at 42°C.

Still another aspect of the present invention is directed to a serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:5 or a nucleotide sequence having at least 50% similarity to all or part thereof or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:5 under low stringency conditions at 42°C.

In another aspect of the present invention, there is provided a serine proteinase in isolated form comprising a sequence of amino acids encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:7 or a nucleotide sequence having at least 50% similarity to all or part thereof or a nucleotide sequence capable of hybridising to the sequence set forth in SEQ ID NO:7 under low stringency conditions at 42°C.

Another embodiment of the present invention is directed to a kinase in isolated form comprising an amino acid sequence substantially as set forth in SEQ ID NO:10 or having 50% amino acid

similarity to all or part thereof. This kinase is referred to herein as "BCON3".

In a related embodiment, the kinase comprises an amino acid sequence encoded by a nucleotide sequence substantially as set forth in SEQ ID NO:9 or a nucleotide sequence having at least 50% similarity to all or part of the nucleotide sequence set forth in SEQ ID NO:9 or a nucleotide sequence capable of hybridising to the nucleotide sequence set forth in SEQ ID NO:9 under low stringency conditions at 42°C.

The present invention further provides an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:3 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:3 under low stringency conditions at 42°C.

Another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:5 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:5 under low stringency conditions at 42°C.

Another aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:7 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:7 under low stringency conditions at 42°C

Still another aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides substantially as set forth in SEQ ID NO:9 or having 50% similarity to all or part thereof or a nucleic acid molecule capable of hybridising to SEQ ID NO:9 under low stringency conditions at 42°C.

Reference herein to a low stringency includes low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for

washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions.

Reference herein to similarity to "part" of a sequence means similarity to at least about 4 contiguous amino acids or at least about 12 contiguous nucleotide bases and more preferably at least about 7 contiguous amino acids or at least about 21 contiguous nucleotide bases.

The term "similarity" includes exact identity between sequences or, where the sequence differs, different amino acids may be related to each other at the structural, functional, biochemical and/or conformational levels.

The term "isolated" includes biological purification and biological separation and encompasses molecules having undergone at least one purification, concentration or separation step relative to its natural environment. For example, a preparation may comprise at least about 10%, preferably at least about 20%, more preferably at least about 30%, still more preferably at least about 50% or greater of the molecule relative to at least one other component in a composition as determined by activity, mass, amino acid content, nucleotide content or other convenient means.

Hereinafter, the molecules of the present invention are referred to as a "proteinase/kinase". The term "proteinase/kinase" includes the serine proteinases HELA2 (testisin) and ATC2 and the kinase BCON3. The proteinase/kinase of the present invention may be in isolated, naturally occurring form or recombinant or synthetic form or chemical analogues thereof.

The proteinase/kinase of the present invention is preferably of human origin but from non-human origins are also encompassed by the present invention. Non-human animals contemplated by the

present invention include primates, livestock animals (e.g. sheep, cows, pigs, goats, horses, donkeys), laboratory test animals (e.g. mice, rats, guinea pigs, hamsters, rabbits), domestic companion animals (e.g. dogs, cats), birds (e.g. chickens, geese, ducks and other poultry birds, game birds, emus, ostriches) and captive wild or tamed animals (e.g. foxes, kangaroos, dingoes). The present invention also encompasses a proteinase/kinase homologue from *Xenopus* and plants.

The nucleic acid molecules encoding a proteinase/kinase may be genomic DNA, cDNA or RNA such as mRNA.

Another aspect of the present invention contemplates a method for cloning a nucleotide sequence encoding a novel serine proteinase, said method comprising screening a nucleic acid library with said one or more or oligonucleotides defined by SEQ ID NO:1 and/or SEQ ID NO:2 and obtaining a clone therefrom which encodes said novel serine proteinase or part thereof.

Preferably, the nucleic acid library is genomic DNA, cDNA, genomic or mRNA library.

Preferably, the nucleic acid library is a cDNA expression library.

Preferably, the nucleic acid library is of human origin such as from brain, liver, kidney, neo-natal tissue, embryonic tissue, tumour or cancer tissue.

With respect to HELA2 (HELA2 (testisin)), significant expression is generally only found in normal testis. Accordingly, the present invention extends to nucleic acid molecules capable of tissue-specific or substantially tissue-specific expression.

Still another embodiment contemplates the promoter or a functional part thereof of the genomic gene encoding the subject proteinase/kinase of the present invention. The promoter may readily be obtained by, for example, "chromosome walking". A particularly useful promoter is from HELA2 (testisin) which can be regarded as a testis specific promoter. This promoter can be used, for example, to direct testis specific expression of genetic sequences operably linked to the

promoter and may be used *inter alia* gene therapy or modulation of fertility.

The present invention further contemplates a range of derivatives of the subject proteinase/kinase. Derivatives include fragments, parts, portions, mutants, homologues and analogues of the subject polypeptides and corresponding genetic sequences. Derivatives also include single or multiple amino acid substitutions, deletions and/or additions to the subject molecules or single or multiple nucleotide substitutions, deletions and/or additions to the genetic sequence encoding the molecules. "Additions" to amino acid sequences or nucleotide sequences include fusions with other peptides, polypeptides or proteins or fusions to nucleotide sequences. Reference herein to the serine proteinase and kinase includes reference to all derivatives thereof including functional derivatives or immunologically interactive derivatives.

Analogues of the subject serine proteinase and kinase contemplated herein include, but are not limited to, modification to side chains, incorporating of unnatural amino acids and/or their derivatives during peptide, polypeptide or protein synthesis and the use of crosslinkers and other methods which impose conformational constraints on the proteinaceous molecule or their analogues.

Examples of side chain modifications contemplated by the present invention include modifications of amino groups such as by reductive alkylation by reaction with an aldehyde followed by reduction with $NaBH_4$; amidination with methylacetimidate; acylation with acetic anhydride; carbamoylation of amino groups with cyanate; trinitrobenzylation of amino groups with 2, 4, 6-trinitrobenzene sulphonic acid (TNBS); acylation of amino groups with succinic anhydride and tetrahydrophthalic anhydride; and pyridoxylation of lysine with pyridoxal-5-phosphate followed by reduction with $NaBH_4$.

The guanidine group of arginine residues may be modified by the formation of heterocyclic condensation products with reagents such as 2,3-butanedione, phenylglyoxal and glyoxal.

The carboxyl group may be modified by carbodiimide activation *via* O-acylisourea formation followed by subsequent derivitisation, for example, to a corresponding amide.

Sulphydryl groups may be modified by methods such as carboxymethylation with iodoacetic acid or iodoacetamide; performic acid oxidation to cysteic acid; formation of a mixed disulphides with other thiol compounds; reaction with maleimide, maleic anhydride or other substituted maleimide; formation of mercurial derivatives using 4-chloromercuribenzoate, 4-chloromercuriphenylsulphonic acid, phenylmercury chloride, 2-chloromercuri-4-nitrophenol and other mercurials; carbamoylation with cyanate at alkaline pH.

Tryptophan residues may be modified by, for example, oxidation with N-bromosuccinimide or alkylation of the indole ring with 2-hydroxy-5-nitrobenzyl bromide or sulphenyl halides. Tyrosine residues on the other hand, may be altered by nitration with tetranitromethane to form a 3-nitrotyrosine derivative.

Modification of the imidazole ring of a histidine residue may be accomplished by alkylation with iodoacetic acid derivatives or N-carbethoxylation with diethylpyrocarbonate.

Examples of incorporating unnatural amino acids and derivatives during peptide synthesis include, but are not limited to, use of norleucine, 4-amino butyric acid, 4-amino-3-hydroxy-5-phenylpentanoic acid, 6-aminohexanoic acid, t-butylglycine, norvaline, phenylglycine, ornithine, sarcosine, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-thienyl alanine and/or D-isomers of amino acids. A list of unnatural amino acid, contemplated herein is shown in Table 3.

TABLE 3

Non-conventional amino acid	Code	Non-conventional amino acid	Code
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-carboxylate	Cpro	L-N-methylasparagine	Nmasn
		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-carboxylate	Norb	L-N-methylglutamine	Nmgln
		L-N-methylglutamic acid	Nmglu
cyclohexylalanine		Chexa L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva

	D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
	D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
	D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
	D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
5	D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
	D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
	D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
	D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
10	D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
	D-α-methylleucine	Dmleu	α-napthylalanine	Anap
	D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
	D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
	D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
15	D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
	D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
20	D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
25	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser
30	D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
	D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp

	D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
	N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
	D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
	N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
	N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
	N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
	N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
	D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
	D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
	D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
	γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
	L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
	L-ethylglycine	Etg	penicillamine	Pen
	L-homophenylalanine	Hphe	L-α-methylalanine	Mala
	L-α-methylarginine	Marg	L-α-methylasparagine	Masn
	L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
	L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
	L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
	L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
	L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
	L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
	L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
	L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
	L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
	L-α-methylserine	Mser	L-α-methylthreonine	Mthr
	L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr

L-α-methylvaline	Mval	L-N-methylhomophenylalanine	Nmhphe
N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe
1-carboxy-1-(2,2-diphenyl-ethylamino)cyclopropane	Nmbc		

Crosslinkers can be used, for example, to stabilise 3D conformations, using homo-bifunctional crosslinkers such as the bifunctional imido esters having $(CH_2)_n$ spacer groups with n=1 to n=6, glutaraldehyde, N-hydroxysuccinimide esters and hetero-bifunctional reagents which usually contain an amino-reactive moiety such as N-hydroxysuccinimide and another group specific-reactive moiety such as maleimido or dithio moiety (SH) or carbodiimide (COOH). In addition, peptides can be conformationally constrained by, for example, incorporation of C_α and N_α-methylamino acids, introduction of double bonds between C_α and C_β atoms of amino acids and the formation of cyclic peptides or analogues by introducing covalent bonds such as forming an amide bond between the N and C termini, between two side chains or between a side chain and the N or C terminus.

These types of modifications may be important to stabilise the proteinase/kinase if administered to an individual or for use as a diagnostic reagent.

The present invention further contemplates chemical analogues of the proteinase/kinase capable of acting as antagonists or agonists of the native molecules or which can act as functional analogues of the native molecules. For example, an antagonist may be a proteinase inhibitor. Chemical analogues may not necessarily be derived from the subject enzymes but may share certain conformational similarities. Alternatively, chemical analogues may be specifically designed to mimic certain physiochemical properties of the serine proteinases or kinases. Chemical analogues may be chemically synthesised or may be detected following, for example, natural product screening.

The identification of the novel molecules of the present invention permits the generation of a range of therapeutic molecules capable of modulating expression of their native counterparts or modulating their activity. Modulators contemplated by the present invention includes agonists and antagonists of proteinase/kinase expression. Antagonists of proteinase/kinase expression include antisense molecules, ribozymes and co-suppression molecules. Agonists include molecules which increase promoter ability or interfere with negative regulatory mechanisms. Agonists of proteinase/kinase include molecules which overcome any negative regulatory mechanism. Antagonists of the proteinase/kinase include antibodies and inhibitor peptide fragments.

Other derivatives contemplated by the present invention include a range of glycosylation variants from a completely unglycosylated molecule to a modified glycosylated molecule. Altered glycosylation patterns may result from expression of recombinant molecules in different host cells.

Another embodiment of the present invention contemplates a method for modulating expression of proteinase/kinase in a human, said method comprising contacting the proteinase/kinase gene encoding proteinase/kinase with an effective amount of a modulator of proteinase/kinase expression for a time and under conditions sufficient to up-regulate or down-regulate or otherwise modulate expression of proteinase/kinase. For example, a nucleic acid molecule encoding proteinase/kinase or a derivative thereof may be introduced into a cell conversely, proteinase/kinase antisense sequences such as oligonucleotides may be introduced.

Another aspect of the present invention contemplates a method of modulating activity of proteinase/kinase in a human, said method comprising administering to said mammal a modulating effective amount of a molecule for a time and under conditions sufficient to increase or decrease proteinase/kinase activity. The molecule may be a proteinaceous molecule or a chemical entity and may also be a derivative of proteinase/kinase or its receptor or a chemical analogue or truncation mutant of proteinase/kinase or its receptor.

One particularly useful serine proteinase, HELA2 (testisin), is implicated in spermatogenesis

and in testicular tumour development. It is proposed, in accordance with the present invention, that HELA2 (testisin) is involved in fertility and infertility.

Northern blot analysis of Poly A+ RNA from normal tissue specimens showed a unique tissue distribution for HELA2 (testisin) with significant expression only in the testis. No signals are detected in any other tissue, with the exception of a minor signal in salivary gland. By RT-PCR, HELA2 (testisin) is detected in the ejaculate of normal males but not in the ejaculate of vasectomised males indicating that it is of germ cell origin. Hybridization data *in situ* indicated that HELA2 (testisin) is produced by immature germ cells in the testis, located near the basal epithelium and, hence, is an important factor for normal sperm maturation; defective expression or mutations would contribute to primary male infertility. Further, it is from the precursors of spermatocytes that 95% of testicular germ cell tumours, such as seminomas, embryonal carcinomas and teratocarcinomas arise. In the normal testis, germ cells undergo meiosis to become spermatocytes, but in individuals at risk, the germ cells continue to proliferate giving rise to germ cell tumours. Although not wishing to limit the present invention to any one theory or mode of action, it is proposed, in accordance with present invention, that HELA2 (testisin) functions at this critical juncture - cell growth versus maturation.

Familial forms of testicular cancer are rare, but linkage analysis of a large family with familial seminoma has demonstrated linkage to chromosome 16p, within a region adjacent to the HPKD1 (human polycystic kidney disease) gene at 16p13.3. The HELA2 (testisin) gene localises to chromosome 16p13.3 which is near the telomere of chromosome 16 and is associated with high genetic instability. The HELA2 (testisin) gene is sandwiched between four genes which underlie other human genetic disorders; HPKD1 and tuberous sclerosis (TSC2) on the one side, and familial mediterranean fever (MEF) and Rubenstein-Taybi syndrome (RSTS) on the other side. The question of whether HELA (testisin) may be a tumour suppressor for seminoma was determined by comparing HELA (testisin) mRNA expression in normal testes with corresponding germ cell tumours from patients with seminoma. HELA (testisin) was not detectable in the tumours of these patients, but was present in the corresponding normal testis specimens, indicative of a tumour suppressor role of HELA2 (testisin) in testicular germ cell cancers.

Although restricted in normal tissues to the testes, HELA2 (testisin) is expressed in tumours of the colon, pancreas, prostate and ovary. This indicates that HELA (testisin) contributed to tumourigenesis and, therefore, has an application as a marker and also as a therapeutic anti-tumour target in these types of cancers.

These data point to a potentially very significant role for HELA2 (testisin) in testicular germ cell maturation (spermatogenesis) as well as in the genesis of testicular germ cell tumours. In accordance with the present invention, it is proposed that expression of HELA2 (testisin) by immature germ cells may be essential for sperm cell development, such that loss of HELA2 (testisin) expression leads to continued and uncontrolled proliferation of immature germ cells leading to subsequent tumourigenesis. Germ cells wherein HELA2 (testisin) is mutated or absent may thus be prone to malignant transformation because of an inability to progress along the differentiation pathway.

HELA2 (testisin) is well-positioned to anchor on the surface of the germ cell where it would participate in a range of proteolytic activities, including cell migration, differentiation and/or activation of growth factors, receptors, or cytokines as well as initiate additional proteolytic cascades. Although not intending to limit the present invention to any one theory or mode of action, it is proposed, in accordance with the present invention, that the proteolytic target of HELA2 (testisin) is a cytokine, receptor or growth factor essential for either germ cell proliferation or differentiation - ie. HELA2 (testisin) may either inactivate a factor important for proliferation, or activate a factor which promotes differentiation. Thus, HELA2 (testisin) may be critical in the regulation of specific cytokines, cytokine receptors or growth factors by means of post-translational proteolytic processing. That HELA2 (testisin) is not present in other normal tissues of the male urogenital tract, such as the prostate and kidney, also argues for such a role specific to the testis.

Diagnostic and therapeutic applications for HELA2 (testisin) have the potential to be wide-ranging both in the cancer and fertility/infertility markets. In tumours, other than the testis, it is desirable to block or inhibit HELA2 (testisin) activity. As HELA2 (testisin) is a member of the serine proteinase family, for which protype crystal structures are known and the catalytic

mechanism reasonably well characterised, the design of drugs that target HELA2 (testisin) proteolytic activity as an anti-tumour therapy should be relatively straightforward. As HELA2 (testisin) is predicted to be anchored on the cell surface, there would not be difficulties associated with delivery of drugs to intracellular compartments. Further, it is very possible that some tumour-associated HELA2 (testisin) may be proteolytically cleaved from the surface of tumour cells, and the extracellular domain detectable in patient serum as a potential tumour associated marker.

Testicular cancer is the commonest malignancy in men aged 20-44 years. Early diagnosis correlates which an improved chance of cure and in a reduction in the severity of treatment. If the cancer is not treated early, it becomes very aggressive. The incidence of testicular cancer is significant (9/100,000) and has been rising over the last 10 years. In testicular germ cell tumours, such as seminoma, delivery of recombinant HELA2 (testisin) using gene therapy techniques could lead to arrest of tumour growth and potentially allow commencement of normal sperm cell maturation and differentiation, thereby reducing the need for surgical removal of the testis (orchidectomy). This may be particularly effective for patients who have already had one testicle removed because of testicular cancer. The risk of contralateral testicular cancer is increased in these patients and tumour development could be arrested through early treatment with HELA2 (testisin) to arrest growth and assist maturation of germ cells. The finding of mutant forms of HELA2 (testisin) may also lead to new markers for seminoma. Unlike other testicular non-seminoma cancers where α-fetoprotein and β-HCG are frequently elevated and can be used as tumour markers, the lack of an adequate marker for seminoma creates difficulties with staging and patient follow-up.

A demonstrated role for HELA2 (testisin) in sperm maturation and development would likely lead to improved diagnosis and new directed therapeutics for male primary infertility. Primary male infertility is responsible for conception problems in 5-10% of couples and the world market for a therapeutic in this area would be very substantial. Delivery of recombinant HELA2 (testisin) could assist sperm maturation and potentially trigger normal sperm development in some of these cases. The identification of mutant forms of HELA2 (testisin) could air in diagnosis of infertility. If HELA2 (testisin) does not prove to be a tumour

suppressor, but is important for sperm maturation, it could provide a new target for the development of a male contraceptive. If hormonal regulation of HELA2 (testisin) can be demonstrated, HELA2 (testisin) may prove effective for the treatment of conditions arising from dysfunctional hormal responses, such as cryptorchidism, which is associated with both infertility and seminoma development.

Accordingly, the present invention contemplates a pharmaceutical composition comprising proteinase/kinase or a derivative thereof or a modulator of proteinase/kinase expression or proteinase/kinase activity and one or more pharmaceutically acceptable carriers and/or diluents. These components are referred to as the "active ingredients" and include, for example, HELA2 (testisin).

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) and sterile powders for the extemporaneous preparation of sterile injectable solutions. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as licithin, by the maintenance of the required particle size in the case of dispersion and by the use of superfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thirmerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion

medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the active ingredients are suitably protected they may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 1% by weight of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 5 to about 80% of the weight of the unit. The amount of active compound in such therapeutically useful compositions in such that a suitable dosage will be obtained. Preferred compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.1 μg and 2000 mg of active compound.

The tablets, troches, pills, capsules and the like may also contain the components as listed hereafter. A binder such as gum, acacia, corn starch or gelatin; excipients such as dicalcium phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid and the like; a lubricant such as magnesium stearate; and a sweetening agent such a sucrose, lactose or saccharin may be added or a flavouring agent such as peppermint, oil of wintergreen, or cherry flavouring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavouring such as cherry or orange flavour. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound(s) may be incorporated into sustained-release

preparations and formulations.

Pharmaceutically acceptable carriers and/or diluents include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Parental compositions are generally suitable for administration by the intraveneous, subcutaneous or intramuscular routes amongst other routes of administration. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly dependent on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such an active material for the treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail. Other forms of administration include but are not limited to intranasal, buccal, rectal, suppository, inhalation, intracerebral and intraperitoneal.

The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 μg to about 2000 mg. Expressed in proportions, the active compound is generally present in from about 0.5 μg to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients.

The effective amounts include amounts calculated or predicted to have the desired effect and range from at least about 0.01 ng/kg body weight to about 10,000 mg/kg body weight. Alternative amounts include 0.1 ng/kg body weight to about 1000 ng/kg body weight.

The pharmaceutical composition may also comprise genetic molecules such as a vector capable of transfecting target cells where the vector carries a nucleic acid molecule capable of modulating proteinase/kinase expression or proteinase/kinase activity. The vector may, for example, be a viral vector. This form of therapy is proposed to be particularly useful for gene replacement or enhancement therapy for HELA2 (testisin) especially for the modulation of fertility and/or treatment of testicular cancer.

Still another aspect of the present invention is directed to antibodies to proteinase/kinase and its derivatives. Such antibodies may be monoclonal or polyclonal and may be selected from naturally occurring antibodies to proteinase/kinase or may be specifically raised to proteinase/kinase or derivatives thereof. In the case of the latter, proteinase/kinase or its derivatives may first need to be associated with a carrier molecule. .The antibodies and/or recombinant proteinase/kinase or its derivatives of the present invention are particularly useful as therapeutic or diagnostic agents. For example, monitoring non-testicular cancer by measuring HELA2 (testisin) or screening for the presence of testicular cancer by an absence of HELA2 (testisin).

Proteinase/kinase and its derivatives may also be used to screen for naturally occurring antibodies to proteinase/kinase. These may occur, for example in some autoimmune diseases. Alternatively, specific antibodies can be used to screen for proteinase/kinase. Techniques for such assays are well known in the art and include, for example, sandwich assays and ELISA. Knowledge of proteinase/kinase levels may be important for diagnosis of certain cancers or a predisposition to cancers or for monitoring certain therapeutic protocols.

Antibodies the proteinase/kinase of the present invention may be monoclonal or polyclonal. Alternatively, fragments of antibodies may be used such as Fab fragments. Furthermore, the present invention extends to recombinant and synthetic antibodies and to antibody hybrids. A

"synthetic antibody" is considered herein to include fragments and hybrids of antibodies. The antibodies of this aspect of the present invention are particularly useful for immunotherapy and may also be used as a diagnostic tool for assessing apoptosis or monitoring the program of a therapeutic regimen.

For example, specific antibodies can be used to screen for proteinase/kinase proteins. The latter would be important, for example, as a means for screening for levels of proteinase/kinase in a cell extract or other biological fluid or purifying proteinase/kinase made by recombinant means from culture supernatant fluid. Techniques for the assays contemplated herein are known in the art and include, for example, sandwich assays and ELISA.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies or synthetic antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody. An antibody as contemplated herein includes any antibody specific to any region of proteinase/kinase.

Both polyclonal and monoclonal antibodies are obtainable by immunization with the enzyme or protein and either type is utilizable for immunoassays. The methods of obtaining both types of sera are well known in the art. Polyclonal sera are less preferred but are relatively easily prepared by injection of a suitable laboratory animal with an effective amount of proteinase/kinase, or antigenic parts thereof, collecting serum from the animal, and isolating specific sera by any of the known immunoadsorbent techniques. Although antibodies produced by this method are utilizable in virtually any type of immunoassay, they are generally less favoured because of the potential heterogeneity of the product.

The use of monoclonal antibodies in an immunoassay is particularly preferred because of the ability to produce them in large quantities and the homogeneity of the product. The preparation of hybridoma cell lines for monoclonal antibody production derived by fusing an immortal cell line and lymphocytes sensitized against the immunogenic preparation can be done by techniques

which are well known to those who are skilled in the art.

Another aspect of the present invention contemplates a method for detecting proteinase/kinase in a biological sample from a subject said method comprising contacting said biological sample with an antibody specific for proteinase/kinase or its derivatives or homologues for a time and under conditions sufficient for an antibody-proteinase/kinase complex to form, and then detecting said complex.

The presence of proteinase/kinase may be accomplished in a number of ways such as by Western blotting and ELISA procedures. A wide range of immunoassay techniques are available as can be seen by reference to US Patent Nos. 4,016,043, 4, 424,279 and 4,018,653. These, of course, includes both single-site and two-site or "sandwich" assays of the non-competitive types, as well as in the traditional competitive binding assays. These assays also include direct binding of a labelled antibody to a target.

Sandwich assays are among the most useful and commonly used assays and are favoured for use in the present invention. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabelled antibody is immobilized on a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen complex, a second antibody specific to the antigen, labelled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of another complex of antibody-antigen-labelled antibody. Any unreacted material is washed away, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labelled antibody are added simultaneously to the bound antibody. These techniques are well known to those skilled in the art, including any minor variations as will be readily apparent. In accordance with the present invention the sample is one which might contain proteinase/kinase including cell extract,

tissue biopsy or possibly serum, saliva, mucosal secretions, lymph, tissue fluid and respiratory fluid. The sample is, therefore, generally a biological sample comprising biological fluid but also extends to fermentation fluid and supernatant fluid such as from a cell culture.

In the typical forward sandwich assay, a first antibody having specificity for the proteinase/kinase or antigenic parts thereof, is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene. The solid supports may be in the form of tubes, beads, discs of microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consist of cross-linking covalently binding or physically adsorbing, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested is then added to the solid phase complex and incubated for a period of time sufficient (e.g. 2-40 minutes or overnight if more convenient) and under suitable conditions (e.g. from about room temperature to about 37°C) to allow binding of any subunit present in the antibody. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of the hapten. The second antibody is linked to a reporter molecule which is used to indicate the binding of the second antibody to the hapten.

An alternative method involves immobilizing the target molecules in the biological sample and then exposing the immobilized target to specific antibody which may or may not be labelled with a reporter molecule. Depending on the amount of target and the strength of the reporter molecule signal, a bound target may be detectable by direct labelling with the antibody. Alternatively, a second labelled antibody, specific to the first antibody is exposed to the target-first antibody complex to form a target-first antibody-second antibody tertiary complex. The complex is detected by the signal emitted by the reporter molecule.

By "reporter molecule" as used in the present specification, is meant a molecule which, by its chemical nature, provides an analytically identifiable signal which allows the detection of antigen-bound antibody. Detection may be either qualitative or quantitative. The most

commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules (i.e. radioisotopes) and chemiluminescent molecules.

In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, beta-galactosidase and alkaline phosphatase, amongst others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable colour change. Examples of suitable enzymes include alkaline phosphatase and peroxidase. It is also possible to employ fluorogenic substrates, which yield a fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labelled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the complex of antibody-antigen-antibody. The substrate will react with the enzyme linked to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample. "Reporter molecule" also extends to use of cell agglutination or inhibition of agglutination such as red blood cells on latex beads, and the like.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labelled antibody adsorbs the light energy, inducing a state to excitability in the molecule, followed by emission of the light at a characteristic colour visually detectable with a light microscope. As in the EIA, the fluorescent labelled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining tertiary complex is then exposed to the light of the appropriate wavelength the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescene and EIA techniques are both very well established in the art and are particularly preferred for the present method. However, other reporter molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed.

The present invention also contemplates genetic assays such as involving PCR analysis to detect proteinase/kinase gene or its derivatives. Alternative methods or methods used in conjunction include direct nucleotide sequencing or mutation scanning such as single stranded conformation polymorphoms analysis (SSCP), specific oligonucleotide hybridisation, and methods such as direct protein truncation tests.

The nucleic acid molecules of the present invention may be DNA or RNA. When the nucleic acid molecule is in DNA form, it may be genomic DNA or cDNA. RNA forms of the nucleic acid molecules of the present invention are generally mRNA.

Although the nucleic acid molecules of the present invention are generally in isolated form, they may be integrated into or ligated to or otherwise fused or associated with other genetic molecules such as vector molecules and in particular expression vector molecules. Vectors and expression vectors are generally capable of replication and, if applicable, expression in one or both of a prokaryotic cell or a eukaryotic cell. Preferably, prokaryotic cells include *E. coli, Bacillus sp* and *Pseudomonas sp*. Preferred eukaryotic cells include yeast, fungal, mammalian and insect cells.

Accordingly, another aspect of the present invention contemplates a genetic construct comprising a vector portion and a mammalian and more particularly a human proteinase/kinase gene portion, which proteinase/kinase gene portion is capable of encoding an proteinase/kinase polypeptide or a functional or immunologically interactive derivative thereof.

Preferably, the proteinase/kinase gene portion of the genetic construct is operably linked to a promoter on the vector such that said promoter is capable of directing expression of said proteinase/kinase gene portion in an appropriate cell.

In addition, the proteinase/kinase gene portion of the genetic construct may comprise all or part of the gene fused to another genetic sequence such as a nucleotide sequence encoding glutathione-S-transferase or part thereof.

The present invention extends to such genetic constructs and to prokaryotic or eukaryotic cells comprising same.

The present invention also extends to any or all derivatives of proteinase/kinase including mutants, part, fragments, portions, homologues and analogues or their encoding genetic sequence including single or multiple nucleotide or amino acid substitutions, additions and/or deletions to the naturally occurring nucleotide or amino acid sequence. The present invention further encompasses hybrids between the proteinase/kinases such as to broaden the spectrum of activity and to ligands and substrates of the proteinase/kinase.

The proteinase/kinase and its genetic sequence of the present invention will be useful in the generation of a range of therapeutic and diagnostic reagents.

Soluble proteinase/kinase polypeptides or other derivatives, agonists or antagonists are also contemplated to be useful in the treatment of disease, injury or abnormality in the nervous system, e.g. in relation to central or peripheral nervous system to treat Cerebral Palsy, trauma induced paralysis, vascular ischaemia associated with stroke, neuronal tumours, motoneurone disease, Parkinson's disease, Huntington's disease, Alzheimer's disease, Multiple Sclerosis, peripheral neuropathies associated with diabetes, heavy metal or alcohol toxicity, renal failure and infectious diseases such as herpes, rubella, measles, chicken pox, HIV or HTLV-1. Other conditions for which the proteinase/kinase are useful include cancer, metastasis and autoimmune disease amongst many others. Particular applications for HELA2 (testisin) include as a marker for non-testicular cancers, in the treatment of testicular cancer and in the treatment of infertility or in inducing infertility such for contraception.

A further aspect of the present invention contemplates the use of proteinase/kinase or its functional derivatives in the manufacture of a medicament for the treatment of proteinase/kinase mediated conditions defective or deficient.

The present invention is further described by the following non-limiting Examples.

EXAMPLE 1
CLONING PROCEDURES

In order to identify serine proteinases that may be involved in regulatory cellular functions, a genetic screening approach was applied using degenerate primers corresponding to conserved regions of serine proteinases (amino acids flanking His- and Ser- residues) to amplify gene fragments spanning these regions from cDNA, using a low stringency RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction) approach.

By this technique, the aim was to isolate low abundance genes as well as those present in moderate to high abundance. The cDNA used for these experiments was isolated from a HeLa cell cytotoxicity model wherein PAI-2 expression inhibits TNFα-induced apoptosis (Dickinson, *et al J. Biol. Chem. 270,* 27894-27904, 1995). These PAI-2 expressing cells provide a unique and viable system for investigating TNF signalling pathways as they are protected from the cytotoxic effects of TNFα.

cDNA was generated from RNA isolated from PAI-2 expressing HeLa cells, untreated and following treatment with TNF and cycloheximide. Amplification of both cDNA populations using PCR and the following serine proteinase degenerate primers,

His Primer: 5'ACAGAATTCTGGGTIGTIACIGCIGCICAYTG3' [SEQ ID NO:1],
Ser Primer: 5'ACACTTAAGAXIGGICCICCIC/GT/AXTCICC3' [SEQ ID NO:2]

produced DNA fragments in the range of 480bp, the approximate predicted size of the serine proteinase intergenic region. These amplified DNA fragments were cloned into *E. coli* generating a library containing approximately 150 independent clones. The inventors analysed 36 of these clones and found that 9 encoded previously identified serine proteinases or tissue-type or urokinase-type plasminogen activators, thereby demonstrating the efficacy of this approach. Of the other 36, two were found to encode novel open reading frames with high homology to serine proteinases and are referred to herein as "HELA2" (or "testisin") and "ATC2". One additional clone designated herein , "BCON3", showed homology to a kinase.

Extension of the DNA fragments by RACE in both 5′ and 3′ directions using internally derived primers has verified their homology to the serine proteinase family. The DNA sequences are unique in that they are markedly different from any known DNA or protein sequence in the Genbank and NBRF databases.

EXAMPLE 2
HELA2 SERINE PROTEINASE

The HELA2 mRNA transcript is approximately 1.5kb as determined from Northern blots. Nucleic acid sequence was obtained for about 1.1kb of HELA2 which spans the entire coding region, the 3′ noncoding region and part of the 5′ noncoding region. The coding region starts with an ATG codon which is present in a motif analogous to the Kozak eukaryotic translation initiation consensus sequence. Alignment of the deduced amino acid sequence of HELA2 with homologous serine proteinases shows that the cDNA encodes a 314 amino acid (aa) polypeptide with a calculated molecular weight of 34.8kD, which is synthesized as a zymogen containing pre-, pro- and catalytic regions. The pro- region (or light chain) and the catalytic region (heavy chain) are delineated by a classic serine proteinase activation motif Arg-Ile-Val-Gly-Gly [SEQ ID NO:24] with cleavage likely occurring between Arg and Ile. The catalytic region includes the catalytic triad of His, Asp and Ser in positions and motifs which are highly conserved among the serine proteinases. Ten Cys residues occur in conserved positions: by analogy to other serine proteinases, eight of these function to form disulfide bridges within the catalytic region and the remaining two link the pro- and catalytic regions.

Structural features conserved in the binding pockets of serine proteinases are present in HELA2. An Asp residue at the bottom of the serine proteinase binding pocket six residues before the active site Ser in HELA2 indicates that HELA2 has trypsin-like specificity, with proteolytic cleavage after Arg or Lys in target substrates. HELA2 also contains a conserved Ser-Trp-Gly motif at the top of the binding pocket which is involved in hydrogen bonding with target substrates in other serine proteinases. Two isoforms of HELA2 were identified in a HeLa cell cDNA library (Stratagene UniZap HeLa Library) which differ by an insertion of 2 aa's (Tyr-Ser) within the catalytic binding pocket. At the DNA level there is a corresponding insertion

of 6 nucleotides which generates a *Sfi*1 restriction enzyme site.

A hydrophobicity plot of the HELA2 amino acid sequence (Figure 1) identifies two hydrophobic regions, one located at the amino terminus and the other at the carboxy terminus. The 20 aa amino terminal hydrophobic region is likely to be a signal peptide, which would direct newly synthesized HELA2 to enter the endoplasmic reticulum. The 16 aa hydrophobic carboxy terminus of HELA2 aligns with the transmembrane domain of prostasin, suggesting that HELA2 is likely to be a membrane-anchored serine proteinase. In prostasin, this protruding carboxy terminus may be cleaved, thus releasing the serine proteinase from the membrane. A unique feature of HELA2 is the presence of a putative mitochondrial localisation sequence directly after the signal peptide sequence.

Preliminary Northern blot data show (Figure 2) that HELA2 mRNA is expressed in the epithelial-like HeLa cell, but is not present in the monocyte-like U937 and MonoMac6 cells, suggesting that HELA2 expression demonstrates a degree of cell-type specificity and therefore is likely to have a specific physiological function.

The HELA2 cDNA was cloned in two isoforms, a short isoform and a long isoform. The nucleotide and corresponding amino acid sequence for the short isoform of HELA2 is shown in SEQ ID NOs. 3 and 4, respectively. The long isoform is shown in SEQ ID NO:5 and 6, respectively.

EXAMPLE 2

HELA2 ENCODES THE SERINE PROTEINASE HELA2 (TESTISIN)

A partial cDNA fragment for HELA2 was isolated by a homology cloning approach and full length HELA2 cDNA was obtained by a combination of library screening, 5′ RACE and use of the EST database (see Example 1). HELA2 recognises an mRNA transcript of approximately 1.5kb and encodes a 314 aa protein with a calculated molecular weight of 34.8 kD, called HELA2 (testisin). Alignment of the deduced amino acid sequence of HELA2 (testisin) with the most homologous known serine proteinase, prostasin, shows that HELA2

(testisin) is a putative zymogen containing pre, pro and catalytic regions (Figure 4). The pro region (or light chain- 32 amino acids) and the catalytic region (heavy chain- 263 amino acids) are delineated by a classic serine proteinase activation motif Arg-Ile-Val-Gly-Gly [SEQ ID NO:24] with cleavage likely occurring between Arg and Ile. The catalytic region includes the catalytic triad of His, Asp and Ser in positions and motifs which are highly conserved among other serine proteinases. Ten Cys residues occur in conserved positions: by analogy to other serine proteinases, eight of these are likely to function to form disulfide bridges within the catalytic region and the remaining two like the pro and catalytic regions.

Additional structural features conserved in the binding pockets of serine proteinases are present in HELA2 (testisin). An Asp residue at the bottom of the serine proteinase binding pocket, six residues before the active site Ser, indicates that HELA2 (testisin) has trypsin-like specificity with proteolytic cleavage after P1-Arg or P1-Lys in target substrates. HELA2 (testisin) also contains a conserved Ser-Trp-Gly motif at the top of the binding pocket which is involved in hydrogen boding with target substrates in other serine proteinases.

A hydrophobicity plot of the HELA2 (testisin) amino acid sequence identifies two hydrophobic regions, one located at the amino terminal and the other at the carboxy terminal. The 20 amino acid amino terminal hydrophobic region is likely to function as a signal peptide, which could direct newly synthesized HELA2 (testisin) to enter the endoplasmic reticulum. The 16 amino acid hydrophobic carboxy terminus of HELA2 (testisin) aligns with the transmembrane domain of prostasin, suggesting that HELA2 (testisin) is likely to be membrane-embedded. In prostasin, this protruding carboxy terminus may be cleaved, thus releasing the serine proteinase from the membrane. A similar release of HELA2 (testisin) may also occur, although the putative cleavage sites are not highly homologous between these two proteinases.

EXAMPLE 3

GENERATION OF FULL LENGTH cDNA ENCODING HELA2 (TESTISIN)

Partial cDNAs of the short and long isoforms of HELA2 (testisin) were obtained using a combination of library screening techniques. Full length cDNA of the two isoforms was then

generated in pBluescriptSK(-) by ligating restriction enzyme-digested fragments of the partial cDNAs. A plasmid map of the two constructs, pBluescriptHELA2(S) and pBluescriptHELA2(L), and a restriction enzyme map of the cDNA of the long isoform are shown in Figure 5.

EXAMPLE 4

EXPRESSION OF RECOMBINANT HELA2 (TESTISIN) IN *E.COLI*

(A) Generation of expression constructs

To reduce potential toxic effects on host cells, and therefore optimise expression, a strategy was employed to eliminate the hydrophobic residues of the secretory and membrane anchoring domains of HELA2 (testisin) (testisin $^{(20\text{-}295)}$). Testisin $^{(20\text{-}295)}$ fragments which were His_6 tagged at either the amino or carboxy terminal were obtained by PCR and expression constructs were generated by inserting these into pQE vectors (Qiagen).

The primers used to generate the amino-terminal tagged protein were:
forward: 5′ GCACAGTCGACCAAGCCGGAGTCGCAGAG 3′ [SEQ ID NO:11] and
reverse: 5′ GCACAAAGCTTGCCAGGAGGGGTCTGGCTG 3′ [SEQ ID NO:12]
The amplification product of 858bp was digested with *Sal*I and *Hin*dIII and ligated into pQE-10 to give pQE-10(20-295)N.

The primers used to generate the carboxy-terminal tagged protein were:
forward: 5′ GCACAACCATGGCCAAGCCGGAGTCGCAGGAG 3′ [SEQ ID NO:13] and
reverse 5′ GCACAAGATCTCCAGGAGGGGTCTGGCTG 3′ [SEQ ID NO:14].
The amplification product of 859 bp was digested with *Nco*I and *Bg*III and ligated into PQE-60 to give pQE-60(20-295)C. The constructs are shown in Figure 6.

(b) Expression in *E. coli*

The plasmids were electro-transformed into *E. coli* DH5α cells. To express recombinant

HELA2 (testisin) protein, transformed cells were grown to log phase then induced for 4 hours in the presence of 2mM IPTG. Cells were lysed in a denaturing lysis buffer containing 8M urea, 0.1M NaH_2PO_4 and 0.01M Tris/HCI pH8. Alternatively the cells were lysed in a non-denaturing lysis buffer containing 0.1M NaH_2PO_4, 0.1M NaCl and 0.01 M Tris/HCI pH8. The His_6 tagged protein was recovered by mixing the lysate with a metal affinity resin (Qiagen or Clontech). Purified testisin(L) was eluted with 100 mM EDTA in lysis buffer (pH 6.3). A single band of approximately 32 kDa was obtained (Figure 7A). Western blot analysis using an anti-His_6 antibody showed that the band at 32kDa was His_6 tagged HELA2 (testisin) (Figure 7B).

EXAMPLE 5
IMMUNOLOGY

(A) Rabbit Polyclonal Antibodies Directed Against HELA2 (testisin) Peptide Antigens

Three peptides (listed below and in Figure 8) were selected from the HELA2 (testisin) amino acid sequence on the basis of predicted antigenicity, hydrophilicity and lack of identity with know proteins. These were synthesized (Auspep) and coupled to keyhole limpet hemocyanin. The coupled peptide (500 μg) in PBS (0.5 ml) was emulsified in an equal volume of Freund's complete adjuvant before injection into a rabbit. Booster injections of coupled peptide in Freund's incomplete adjuvant were made at intervals of 2 to 3 weeks. Each rabbit was bled (approximately 1 ml) before the initial injection and about 7 days after the second and subsequent boosters so that the antibody titre could be assessed by ELISA. When a sufficiently high titre was achieved (after 3 to 5 boosters) between 12 and 25 ml of blood was taken from the animal.

Peptide antigen T20-33	KPESQEAAPLSGPC [SEQ ID NO:15]
Peptide antigen T46-63	EDAELGRWPWQGSLRLWDC [SEQ ID NO:16]
Peptide antigen T175-190	GYIKEDEALPSPHTLQC [SEQ ID NO:17]

Antisera showing immunogenicity against each of the peptide antigens was obtained as

demonstrated by direct ELISA assay (Figure 9). These antibodies are then characterised for immunoreactivity against recombinant HELA2 (testisin) and HELA2 (testisin) in cell extracts.

(B) Rabbit Polyclonal Antibodies Directed Against Purified Bacterially Expressed HELA2 (Testisin)

An SDS-PAGE gel slice containing purified His_6 tagged HELA2 (testisin) (see Example 4, part (b)) is to be combined with adjuvant and rabbits immunized as described above. Rabbit antisera are tested by Western blot analysis for immunoreactivity against purified recombinant HELA2 (testisin) and HELA2 (testisin) in cell extracts, as well as use in immunohistochemical analyses.

EXAMPLE 6
EXPRESSION OF HELA2 (TESTISIN) IN EUKARYOTIC CELLS

(A) Generation of expression constructs

Eukaryotic expression constructs encoding testisin(s) and testisin(L) His_6 tagged at the carboxy terminal were generated in the eukaryotic expression vector pcDNA3 (Invitrogen). DNA fragments encoding HELA2 (testisin) were generated by PCR from both pBluescriptHELA2(S) and pBluescriptHELA2(L) using the primers:
forward: 5′ GCACAGGTACCGAGGCCATGGGCGCGCGC 3′ [SEQ ID NO:18] and
reverse 5′ GCACATCTAGATCAGTGGTGGTGGTGGTGGACCGGCCCCAGGA GTGG 3′ [SEQ ID NO:19]

The PCR product of 985 bp obtained from amplification of HELA2 (testisin) from pBluescriptHELA2(S) as template was ligated into pGEM-T (Easy) vector (Promega). Digestion of this shuttle construct with *Not*I released a 1025 bp fragment which was ligated into pcDNA3 generating the short isoform expression construct pcDNA3-Test(S-C). PCR amplification of the long isoform template gave a 991 bp product which was ligated into pGEM-T (easy) vector. *Not*I digestion of the shuttle construct released a 1031 bp fragment which was ligated into pcDNA3 giving pcDNA3-Test(L-C).

Soluble testisin $^{(1-295)}$-His_6 in which the membrane anchoring sequence is deleted and the protein is carboxy-His_6 tagged is to be obtained by PCR amplification of HELA2 (testisin) from pBluescriptHELA2(L) using the primers:

forward: 5′ GCACAGCGGCCGCGAGGCCATGGGCGCGCGC 3′ [SEQ ID NO:20] and

reverse: 5′ GCACAGCGGCCGCTCAGTGGTGGTGGTGGTGGTGCCAGGAGGGGTC TGGCTG 3′ [SEQ ID NO:21].

The PCR product will be digested with *Not*I and ligated into pcDNA3 generating the long isoform expression construct pcDNA3-Test(1-295)L-C. These constructs are detailed in Figure 10.

(B) Expression and cellular localisation of HELA2 (testisin)

Each of the expression constructs is transiently transfected into a eukaryotic cell line (eg. HeLa, CHO or COS cells) by electroporation. Expression is confirmed by Northern blot and immunoblot. The His_6 tag is a small, uncharged tag which reportedly does not interfere with cellular membrane interactions and is able to be detected with anti-His_6 antibodies. HELA2 (testisin) cellular localisation is analysed by immunofluorescence using antibodies directed against the His_6 tag and stained cells examined by confocal microscopy. Mock transfected cells is monitored as one of the controls in these experiments. Cells are examined under non-permeablised and permeabilised conditions to investigate intracellular and cell surface expression of HELA2 (testisin) tagged proteins. Possible release of HELA2 (testisin) into the supernatant is monitored by immunoblotting of conditioned media. Association of HELA2 (testisin) with a particular cellular compartment is confirmed by cellular fractionation studies. Stable transfectants of the full length and truncated tagged HELA2 (testisin) is generated by selection in G418. Recombinant HELA2 (testisin) is purified from these stable transfectants using a metal affinity resin (eg. Qiagen or Clontech) for assay of its bioactivity and efficacy as a therapeutic reagent.

EXAMPLE 7

HELA2 (TESTISIN) IS SPECIFICALLY EXPRESSED IN THE NORMAL TESTIS, AND IS ASSOCIATED WITH SPERM DEVELOPMENT

(A) Normal Tissue Blot

Northern blot analysis of PolyA RNA from 50 normal tissue specimens (strandardised to 8 different housekeeping genes) (Clontech) with a ^{32}P-labelled HELA2 (testisin) probe showed high level expression of HELA2 (testisin) only in the testis (Figure 11A). Hybridization of the radiolabelled probe was in ExpressHyb™ solution (Clontech) at 65°. The blots were washed to a final stringency of 0.1xSSC/0.5% w/v SDS. As a control Figure 11B shows a Northern blot of a second cDNA, BCON3, which is ubiquitously expressed.

(B) HELA2 (testisin) is Expressed in Sperm Cells, Demonstrating its Germ Cell Origin

To determine whether HELA2 (testisin) expression is associated with germ cells of the testis, ejaculate specimens from normal fertile males were compared with those of post-vasectomy males by RT-PCR analysis using HELA2 (testisin) specific primers. Sperm is the only product form the testis that is found in ejaculate; other components of the ejaculate are derived from the prostate.

First strand cDNA reverse transcribed from total RNA which has been isolated from frozen or hanks ejaculate specimens. PCR was performed on 1 μl of cDNA using the primers:
forward: 5′ CTGACTTCCATGCCATCCTT 3′ [SEQ ID NO:22] and
reverse: 5′ GCTCACGACTCCAATCTGAT 3′ [SEQ ID NO:23].
As shown in Figure 12, strong signals of the expected size of 464 bp were detected in ejaculate from normal males (Patients #23 and #31), while no HELA2 (testisin) was detected in Patient #153 (post-vasectomy). Patent #90 (post-vasectomy) showed a low level of amplification product which may reflect a small amount of residual sperm in the seminiferous tubules. PCR using primers specific for β2-macroglobulin was performed on the same samples as a control for the presence of approximately equal amounts of cDNA in each sample.

(C) HELA2 (testisin) is Expressed in Immature Germ Cells of the Testis

In situ hybridization was performed on paraffin-embedded specimens of rat and human testis tissues. The results showed that HELA2 (testisin) mRNA was not associated with the Leydig cells and the pattern was not typical for Sertoli cell staining. Positive staining was detected near the basal almina of the seminiferous tubule in a region associated with spermatocytes and spermatogonia. The presence of HELA2 (testisin) mRNA in these cells indicates a role for HELA2 (testisin) in germ cell maturation and sperm development.

A high level of amino acid identity between HELA2 (testisin) and prostasin at the carboxy terminal indicates that HELA2 (testisin) may anchor on the germ cell surface where it could participate in a range of proteolytic activities, including participation in cell migration, differentiation and/or activation of growth factors and proteolytic cascades. The fact that HELA2 (testisin) is not present in other tissues of the male urogenital tract, such as the prostate and kidney, also argues for such a role specific to the testis.

EXAMPLE 8

HELA2 (TESTISIN) EXPRESSION IS ASSOCIATED WITH TUMOUR IN NON-TESTIS CELL-TYPES

The tissue and cell-type distribution of testisin mRNA transcripts were determined by Northern hybridization analyses of mRNA extracted from cells lines derived from different cancerous tissues (breast, colon, melanoma, monocytic). A summary of these is given in Figure 13. Matches of the HELA2 (testisin) cDNA sequence with EST database entries indicate that HELA2 (testisin) is also present in tumour tissues of the colon, pancreas, prostate and ovary. The presence of HELA2 (testisin) in tumours where it is not expressed normally indicates that it may play a role in tumourigenesis in these cell-types.

EXAMPLE 9

THE HELA2 (TESTISIN) GENE IS LOCATED ON HUMAN CHROMOSOME 16p13.3

The genetic location of testisin was mapped by fluorescence *in-situ* hybridization to normal metaphase chromosomes to the short arm of chromosome 16 at 16p13.3 (Figure 14). Screening of a chromosome 16 hybrid panel then sublocalised *testisin* to cosmids 399A1, 406D6, 54G6 which have been mapped to this region (Figure 15) (Sood *et al* (1997). *Genomics 42:* 83-95). These cosmids lie between the markers D16S291 and D16S63 and the gene is located just centromeric to D16S246. This region of the human genome is associated with high genetic instability and telomeric rearrangements underlie a variety of common human genetic disorders. Testisin is sandwiched between the human disease genes PKD1 (polycystic kidney disease) and tuberous sclerosis (TSC2) on the on side, and MEF (familial mediterranean fever) and Rubenstein-Taybi syndrome (RSTS) on the other side as diagrammed in Figure 16.

EXAMPLE 10

HELA2 (TESTISIN) mRNA EXPRESSION IS ABSENT IN TESTICULAR GERM CELL

To determine whether HELA2 (testisin) may play a role in testicular tumourigenesis, HELA2 (testisin) mRNA expression in normal testes and testicular tumour tissue obtained from 4 patients diagnosed with seminoma were compared by Northern blot analysis. HELA2 (testisin) mRNA was detected in normal testes from all four patients but was not present in the corresponding tumours (Figure 17). This is further evidence of a tumour suppressor role for HELA2 (testisin) in germ cell tumours.

EXAMPLE 11

GENOMIC ORGANISATION OF THE HELA2 (TESTISIN) GENE

The HELA2 (testisin) gene is further characterised by determination of its genomic organisation. Intron-extron boundaries were predicted from the genomic organisation of the homologous serine proteinase prostasin. Introns D and E were amplified with exonic oligonucleotides, the sizes of the amplified products determined by gel electrophoresis, and the intron-exon boundary sequences determined by DNA sequencing. The sequence of intron C was determined from clones which were isolated during cloning of the long and short isoforms of HELA2 (testisin) (See Example 12). The intron-exon boundaries and intron sequence of introns A and B is determined. This has led to the genomic map of HELA2 (testisin) as diagrammed in Figure 15. This analysis extends to tumour tissues to ascertain the tole of HELA2 (testisin) as a tumour suppressor.

EXAMPLE 12

THE HELA2 (TESTISIN) SHORT AND LONG ISOFORMS ARE GENERATED BY ALTERNATIVE mRNA SPLICING

Two isoforms of HELA2 (testisin) where identified which differ by an insertion of 2 amino acids (Tyr-Ser) between the catalytic His and Asp residues. These constitute the long (L) and short (s) isoforms. At the DNA level there is a corresponding insertion of 6 nucleotides which generates a *Sfc*1 restriction enzyme site. PCR amplification from single strand cDNA generated from HeLa cell total RNA followed by DNA sequence analysis of the amplified product demonstrated that the two isoforms are generated through the use of two alternative mRNA splice sites. The DNA sequence for the intron and the flanking exons are shown in Figure 19. The resulting insertion of amino acids YS occurs 4 amino acids after the catalytic His residue of HELA2 (testisin). Preliminary molecular modelling suggests the presence of this insertion is likely to alter the catalytic activity of HELA2 (testisin).

EXAMPLE 13

MUTATION ANALYSIS-HELA2 (TESTISIN) AS A TUMOUR SUPPRESSOR

Intronic DNA sequence information generated above (see Example 11) is used to generate primers to amplify HELA2 (testisin) exons for SSCP analyses. Some of these primers are given in Figure 20. Genomic DNA isolated from seminomas and corresponding normal testis as well as genomic DNA from wild-type and affected seminoma family members are analysed by SSCP for altered expression patterns indicative of genetic mutations. Evidence of genetic mutations are also being determined by DNA sequencing.

EXAMPLE 14

MOUSE HELA2 (TESTISIN) IS HOMOLOGOUS TO HUMAN HELA2 (TESTISIN)

A mouse testis EST was identified which encodes the putative mouse homologue of HELA2 (testisin). The cDNA sequence and corresponding amino acid sequence is given in Figure 21. Over the region of the mouse sequence that we have available, the mouse and human sequences are 72% homologous at the DNA level and 63% homologous at the amino acid level.

EXAMPLE 15

ATC2 SERINE PROTEINASE

ATC2 was isolated from the cDNA of PAI-2 expressing HeLa cells following treatment with TNF and cycloheximide. A partial DNA sequence for ATC2 cDNA has been obtained which encompasses the sequence encoding the serine proteinase catalytic region. Additional clones extending to both 5′ and 3′ directions have been obtained. The available nucleic acid sequence of ATC2 cDNA and its deduced amino acid sequence shows that it is a member of the serine proteinase family with homology to hepsin, prostasin, and acrosin. It thus may belong to the same family as HELA2. The catalytic region includes the His, Asp and Ser conserved motifs. Preliminary Northern blot experiments have failed to detect ATC2 mRNA in total RNA isolated from resting HeLa cells, suggesting that it may not be expressed in abundance in these cells,

which may mean that it is tightly regulated. As it was isolated from cells following treatment with TNF and cycloheximide, its expression may be induced by these agents in HeLa cells. These data have potential significance for a role for ATC2 in apoptosis and cell death. ATC2 may be intracellular, extracellular or found on the cell surface and is likely to be involved in regulating cell functions. Thus ATC2 may have potential significance in the treatment of cancer and diseases involving dysregulation of cell growth and survival. The nucleotide and corresponding amino acid sequence of ATC2 is shown in SEQ ID NOs: 7 and 8, respectively.

EXAMPLE 16

BCON3

The deduced amino acid sequence of BCON3 (SEQ ID NO:10) reveals that it is novel. At both the DNA and protein level, BCON3 shows homology to members of the kinase family of proteins. Although it cannot be classified as a member of any particular sub-family of kinases, alignments of the BCON3 protein with the conserved domains of thymidine kinases and tyrosine and serine/threonine protein kinases indicates possible ATP/GTP binding and phosphate transfer regions. Thus, it may be the first member of a new family of kinases. Analysis of the translation product using hydrophobicity plots and the Prosite protein analysis algorithms indicates BCON3 may lack an N-terminal signal sequence (that is, it is likely to encode an intracellular protein) and it possess a nuclear localization signal. BCON3 mRNA is approximately 2300 nucleotides in length. cDNA sequence (SEQ ID NO:9) has been obtained covering about 95% of the transcript and including the 3′ polyA tail. BCON3 mRNA is expressed in normal HeLa cells, as well as in HeLa cell clones that express PAI-2 (Figure 3).

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all

combinations of any two or more of said steps or features.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: THE COUNCIL OF THE QUEENSLAND INSTITUTE OF MEDICAL RESEARCH

(ii) TITLE OF INVENTION: NOVEL MOLECULES

(iii) NUMBER OF SEQUENCES: 24

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: DAVIES COLLISON CAVE
(B) STREET: 1 LITTLE COLLINS STREET
(C) CITY: MELBOURNE
(D) STATE: VICTORIA
(E) COUNTRY: AUSTRALIA
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: AUSTRALIAN PROVISIONAL
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: PO5101/97
(B) FILING DATE: 13-FEB-1997
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: HUGHES, DR E JOHN L
(C) REFERENCE/DOCKET NUMBER: EJH/AF

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: +61 3 9254 2777
(B) TELEFAX: +61 3 9254 2770
(C) TELEX: AA 31787

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACAGAATTCT GGGTIGTIAC IGCIGCICAY TG 32

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1094 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ACACTTAAGA XIGGICCICC IC/GT/AXTCICC 29

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1094 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 17..965

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CGCGGGAGAG GAGGCC ATG GGC GCG CGC GGG GCG CTG CTG CTG GCG CTG
 49
                  Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu
                    1               5                  10

CTG CTG GCT CGG GCT GGA CTC AGG AAG CCG GAG TCG CAG GAG GCG GCG
 97
Leu Leu Ala Arg Ala Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala
             15                  20                  25

CCG TTA TCA GGA CCA TGC GGC CGA CGG GTC ATC ACG TCG CGC ATC GTG
145
Pro Leu Ser Gly Pro Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val
         30                  35                  40

GGT GGA GAG GAC GCC GAA CTC GGG CGT TGG CCG TGG CAG GGG AGC CTG
193
Gly Gly Glu Asp Ala Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu
     45                  50                  55

CGC CTG TGG GAT TCC CAC GTA TGC GGA GTG AGC CTG CTC AGC CAC CGC
241
Arg Leu Trp Asp Ser His Val Cys Gly Val Ser Leu Leu Ser His Arg
 60                  65                  70                  75

TGG GCA CTC ACG GCG GCG CAC TGC TTT GAA ACT GAC CTT AGT GAT CCC
289
Trp Ala Leu Thr Ala Ala His Cys Phe Glu Thr Asp Leu Ser Asp Pro
                 80                  85                  90

TCC GGG TGG ATG GTC CAG TTT GGC CAG CTG ACT TCC ATG CCA TCC TTC
337
Ser Gly Trp Met Val Gln Phe Gly Gln Leu Thr Ser Met Pro Ser Phe
             95                  100                 105

TGG AGC CTG CAG GCC TAC TAC ACC CGT TAC TTC GTA TCG AAT ATC TAT
385
Trp Ser Leu Gln Ala Tyr Tyr Thr Arg Tyr Phe Val Ser Asn Ile Tyr
        110                 115                 120

CTG AGC CCT CGC TAC CTG GGG AAT TCA CCC TAT GAC ATT GCC TTG GTG
433
Leu Ser Pro Arg Tyr Leu Gly Asn Ser Pro Tyr Asp Ile Ala Leu Val
```

```
    125                 130                 135

AAG CTG TCT GCA CCT GTC ACC TAC ACT AAA CAC ATC CAG CCC ATC TGT
481
Lys Leu Ser Ala Pro Val Thr Tyr Thr Lys His Ile Gln Pro Ile Cys
140                 145                 150                 155

CTC CAG GCC TCC ACA TTT GAG TTT GAG AAC CGG ACA GAC TGC TGG GTG
529
Leu Gln Ala Ser Thr Phe Glu Phe Glu Asn Arg Thr Asp Cys Trp Val
                160                 165                 170

ACT GGC TGG GGG TAC ATC AAA GAG GAT GAG GCA CTG CCA TCT CCC CAC
577
Thr Gly Trp Gly Tyr Ile Lys Glu Asp Glu Ala Leu Pro Ser Pro His
            175                 180                 185

ACC CTC CAG GAA GTT CAG GTC GCC ATC ATA AAC AAC TCT ATG TGC AAC
625
Thr Leu Gln Glu Val Gln Val Ala Ile Ile Asn Asn Ser Met Cys Asn
        190                 195                 200

CAC CTC TTC CTC AAG TAC AGT TTC CGC AAG GAC ATC TTT GGA GAC ATG
673
His Leu Phe Leu Lys Tyr Ser Phe Arg Lys Asp Ile Phe Gly Asp Met
    205                 210                 215

GTT TGT GCT GGC AAT GCC CAA GGC GGG AAG GAT GCC TGC TTC GGT GAC
721
Val Cys Ala Gly Asn Ala Gln Gly Gly Lys Asp Ala Cys Phe Gly Asp
220                 225                 230                 235

TCA GGT GGA CCC TTG GCC TGT AAC AAG GAT GGA CTG TGG TAT CAG ATT
769
Ser Gly Gly Pro Leu Ala Cys Asn Lys Asp Gly Leu Trp Tyr Gln Ile
                240                 245                 250

GGA GTC GTG AGC TGG GGA GTG GGC TGT GGT CGG CCC AAT CGG CCC GGT
817
Gly Val Val Ser Trp Gly Val Gly Cys Gly Arg Pro Asn Arg Pro Gly
            255                 260                 265

GTC TAC ACC AAT ATC AGC CAC CAC TTT GAG TGG ATC CAG AAG CTG ATG
865
Val Tyr Thr Asn Ile Ser His His Phe Glu Trp Ile Gln Lys Leu Met
```

```
        270                 275                 280

GCC CAG AGT GGC ATG TCC CAG CCA GAC CCC TCC TGG CCG CTA CTC TTT
913
Ala Gln Ser Gly Met Ser Gln Pro Asp Pro Ser Trp Pro Leu Leu Phe
    285                 290                 295

TTC CCT CTT CTC TGG GCT CTC CCA CTC CTG GGG CCG GTC TGA
   961
Phe Pro Leu Leu Trp Ala Leu Pro Leu Leu Gly Pro Val  *
300                 305                 310

GCCTACCTGA GCCCATGCAG CCTGGGGCCA CTGCCAAGTC AGGCCCTGGT TCTCTTCTGT
1015

CTTGTTTGGT AATAAACACA TTCCAGTTGA TGCCTTGCAG GGCATTTTTC AAAAAAAAAA
1075

AAAAAAAAAA AAAAAAAAA
1094
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 313 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu Leu Leu Ala Arg Ala
 1               5                  10                  15

Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala Pro Leu Ser Gly Pro
            20                  25                  30

Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val Gly Gly Glu Asp Ala
        35                  40                  45

Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu Arg Leu Trp Asp Ser
    50                  55                  60
```

```
His Val Cys Gly Val Ser Leu Leu Ser His Arg Trp Ala Leu Thr Ala
 65                  70                  75                  80

Ala His Cys Phe Glu Thr Asp Leu Ser Asp Pro Ser Gly Trp Met Val
                85                  90                  95

Gln Phe Gly Gln Leu Thr Ser Met Pro Ser Phe Trp Ser Leu Gln Ala
            100                 105                 110

Tyr Tyr Thr Arg Tyr Phe Val Ser Asn Ile Tyr Leu Ser Pro Arg Tyr
        115                 120                 125

Leu Gly Asn Ser Pro Tyr Asp Ile Ala Leu Val Lys Leu Ser Ala Pro
    130                 135                 140

Val Thr Tyr Thr Lys His Ile Gln Pro Ile Cys Leu Gln Ala Ser Thr
145                 150                 155                 160

Phe Glu Phe Glu Asn Arg Thr Asp Cys Trp Val Thr Gly Trp Gly Tyr
                165                 170                 175

Ile Lys Glu Asp Glu Ala Leu Pro Ser Pro His Thr Leu Gln Glu Val
            180                 185                 190

Gln Val Ala Ile Ile Asn Asn Ser Met Cys Asn His Leu Phe Leu Lys
        195                 200                 205

Tyr Ser Phe Arg Lys Asp Ile Phe Gly Asp Met Val Cys Ala Gly Asn
    210                 215                 220

Ala Gln Gly Gly Lys Asp Ala Cys Phe Gly Asp Ser Gly Gly Pro Leu
225                 230                 235                 240

Ala Cys Asn Lys Asp Gly Leu Trp Tyr Gln Ile Gly Val Val Ser Trp
                245                 250                 255

Gly Val Gly Cys Gly Arg Pro Asn Arg Pro Gly Val Tyr Thr Asn Ile
            260                 265                 270

Ser His His Phe Glu Trp Ile Gln Lys Leu Met Ala Gln Ser Gly Met
        275                 280                 285

Ser Gln Pro Asp Pro Ser Trp Pro Leu Leu Phe Phe Pro Leu Leu Trp
    290                 295                 300
```

```
Ala Leu Pro Leu Leu Gly Pro Val  *
305                 310
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1100 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 17..961

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CGCGGGAGAG GAGGCC ATG GGC GCG CGC GGG GCG CTG CTG CTG GCG CTG
 49
                  Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu
                    1               5                  10

CTG CTG GCT CGG GCT GGA CTC AGG AAG CCG GAG TCG CAG GAG GCG GCG
 97
Leu Leu Ala Arg Ala Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala
             15                  20                  25

CCG TTA TCA GGA CCA TGC GGC CGA CGG GTC ATC ACG TCG CGC ATC GTG
145
Pro Leu Ser Gly Pro Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val
         30                  35                  40

GGT GGA GAG GAC GCC GAA CTC GGG CGT TGG CCG TGG CAG GGG AGC CTG
193
Gly Gly Glu Asp Ala Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu
     45                  50                  55

CGC CTG TGG GAT TCC CAC GTA TGC GGA GTG AGC CTG CTC AGC CAC CGC
241
Arg Leu Trp Asp Ser His Val Cys Gly Val Ser Leu Leu Ser His Arg
```

```
 60                  65                  70                  75

TGG GCA CTC ACG GCG GCG CAC TGC TTT GAA ACC TAT AGT GAC CTT AGT
289
Trp Ala Leu Thr Ala Ala His Cys Phe Glu Thr Tyr Ser Asp Leu Ser
                 80                  85                  90

GAT CCC TCC GGG TGG ATG GTC CAG TTT GGC CAG CTG ACT TCC ATG CCA
337
Asp Pro Ser Gly Trp Met Val Gln Phe Gly Gln Leu Thr Ser Met Pro
             95                 100                 105

TCC TTC TGG AGC CTG CAG GCC TAC TAC ACC CGT TAC TTC GTA TCG AAT
385
Ser Phe Trp Ser Leu Gln Ala Tyr Tyr Thr Arg Tyr Phe Val Ser Asn
        110                 115                 120

ATC TAT CTG AGC CCT CGC TAC CTG GGG AAT TCA CCC TAT GAC ATT GCC
433
Ile Tyr Leu Ser Pro Arg Tyr Leu Gly Asn Ser Pro Tyr Asp Ile Ala
    125                 130                 135

TTG GTG AAG CTG TCT GCA CCT GTC ACC TAC ACT AAA CAC ATC CAG CCC
481
Leu Val Lys Leu Ser Ala Pro Val Thr Tyr Thr Lys His Ile Gln Pro
140                 145                 150                 155

ATC TGT CTC CAG GCC TCC ACA TTT GAG TTT GAG AAC CGG ACA GAC TGC
529
Ile Cys Leu Gln Ala Ser Thr Phe Glu Phe Glu Asn Arg Thr Asp Cys
                160                 165                 170

TGG GTG ACT GGC TGG GGG TAC ATC AAA GAG GAT GAG GCA CTG CCA TCT
577
Trp Val Thr Gly Trp Gly Tyr Ile Lys Glu Asp Glu Ala Leu Pro Ser
            175                 180                 185

CCC CAC ACC CTC CAG GAA GTT CAG GTC GCC ATC ATA AAC AAC TCT ATG
625
Pro His Thr Leu Gln Glu Val Gln Val Ala Ile Ile Asn Asn Ser Met
        190                 195                 200

TGC AAC CAC CTC TTC CTC AAG TAC AGT TTC CGC AAG GAC ATC TTT GGA
673
Cys Asn His Leu Phe Leu Lys Tyr Ser Phe Arg Lys Asp Ile Phe Gly
```

```
    205                 210                 215

GAC ATG GTT TGT GCT GGC AAT GCC CAA GGC GGG AAG GAT GCC TGC TTC
721
Asp Met Val Cys Ala Gly Asn Ala Gln Gly Gly Lys Asp Ala Cys Phe
220                 225                 230                 235

GGT GAC TCA GGT GGA CCC TTG GCC TGT AAC AAG GAT GGA CTG TGG TAT
769
Gly Asp Ser Gly Gly Pro Leu Ala Cys Asn Lys Asp Gly Leu Trp Tyr
                240                 245                 250

CAG ATT GGA GTC GTG AGC TGG GGA GTG GGC TGT GGT CGG CCC AAT CGG
817
Gln Ile Gly Val Val Ser Trp Gly Val Gly Cys Gly Arg Pro Asn Arg
            255                 260                 265

CCC GGT GTC TAC ACC AAT ATC AGC CAC CAC TTT GAG TGG ATC CAG AAG
865
Pro Gly Val Tyr Thr Asn Ile Ser His His Phe Glu Trp Ile Gln Lys
        270                 275                 280

CTG ATG GCC CAG AGT GGC ATG TCC CAG CCA GAC CCC TCC TGG CCG CTA
913
Leu Met Ala Gln Ser Gly Met Ser Gln Pro Asp Pro Ser Trp Pro Leu
    285                 290                 295

CTC TTT TTC CCT CTT CTC TGG GCT CTC CCA CTC CTG GGG CCG GTC TGAGCCTACC
  968
Leu Phe Phe Pro Leu Leu Trp Ala Leu Pro Leu Leu Gly Pro Val

300                 305                 310                 315


TGAGCCCATG CAGCCTGGGG CCACTGCCAA GTCAGGCCCT GGTTCTCTTC TGTCTTGTTT
1028

GGTAATAAAC ACATTCCAGT TGATGCCTTG CAGGGCATTT TTCAAAAAAA AAAAAAAAAA
1088

AAAAAAAAAA AA
1100
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 314 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Gly Ala Arg Gly Ala Leu Leu Leu Ala Leu Leu Leu Ala Arg Ala
  1               5                  10                  15

Gly Leu Arg Lys Pro Glu Ser Gln Glu Ala Ala Pro Leu Ser Gly Pro
             20                  25                  30

Cys Gly Arg Arg Val Ile Thr Ser Arg Ile Val Gly Gly Glu Asp Ala
         35                  40                  45

Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu Arg Leu Trp Asp Ser
     50                  55                  60

His Val Cys Gly Val Ser Leu Leu Ser His Arg Trp Ala Leu Thr Ala
 65                  70                  75                  80

Ala His Cys Phe Glu Thr Tyr Ser Asp Leu Ser Asp Pro Ser Gly Trp
                 85                  90                  95

Met Val Gln Phe Gly Gln Leu Thr Ser Met Pro Ser Phe Trp Ser Leu
            100                 105                 110

Gln Ala Tyr Tyr Thr Arg Tyr Phe Val Ser Asn Ile Tyr Leu Ser Pro
        115                 120                 125

Arg Tyr Leu Gly Asn Ser Pro Tyr Asp Ile Ala Leu Val Lys Leu Ser
    130                 135                 140

Ala Pro Val Thr Tyr Thr Lys His Ile Gln Pro Ile Cys Leu Gln Ala
145                 150                 155                 160

Ser Thr Phe Glu Phe Glu Asn Arg Thr Asp Cys Trp Val Thr Gly Trp
                165                 170                 175

Gly Tyr Ile Lys Glu Asp Glu Ala Leu Pro Ser Pro His Thr Leu Gln
            180                 185                 190
```

```
Glu Val Gln Val Ala Ile Ile Asn Asn Ser Met Cys Asn His Leu Phe
        195                 200                 205

Leu Lys Tyr Ser Phe Arg Lys Asp Ile Phe Gly Asp Met Val Cys Ala
    210                 215                 220

Gly Asn Ala Gln Gly Gly Lys Asp Ala Cys Phe Gly Asp Ser Gly Gly
225                 230                 235                 240

Pro Leu Ala Cys Asn Lys Asp Gly Leu Trp Tyr Gln Ile Gly Val Val
                245                 250                 255

Ser Trp Gly Val Gly Cys Gly Arg Pro Asn Arg Pro Gly Val Tyr Thr
            260                 265                 270

Asn Ile Ser His His Phe Glu Trp Ile Gln Lys Leu Met Ala Gln Ser
        275                 280                 285

Gly Met Ser Gln Pro Asp Pro Ser Trp Pro Leu Leu Phe Phe Pro Leu
    290                 295                 300

Leu Trp Ala Leu Pro Leu Leu Gly Pro Val
305                 310
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 799 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 24..799

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
AGTTCAGATG AATGGGACTG TGA GAA CCA TCT GTG ACC AAA TTG ATA CAG
 50
```

```
                                Glu Pro Ser Val Thr Lys Leu Ile Gln
                                  1               5

GAA CAG GAG AAA GAG CCG CGG TGG CTG ACA TTA CAC TCC AAC TGG GAG
 98
Glu Gln Glu Lys Glu Pro Arg Trp Leu Thr Leu His Ser Asn Trp Glu
 10                  15                  20                  25

AGC CTC AAT GGG ACC ACT TTA CAT GAA CTT GTA GTA AAT GGG CAG TCT
146
Ser Leu Asn Gly Thr Thr Leu His Glu Leu Val Val Asn Gly Gln Ser
                 30                  35                  40

TGT GAG AGC AGA AGT AAA ATT TCT CTT CTG TGT ACT AAA CAA GAC TGT
194
Cys Glu Ser Arg Ser Lys Ile Ser Leu Leu Cys Thr Lys Gln Asp Cys
             45                  50                  55

GGG CGC CGC CCT GCT GCC CGA ATG AAC AAA AGG ATC CTT GGA GGT CGG
242
Gly Arg Arg Pro Ala Ala Arg Met Asn Lys Arg Ile Leu Gly Gly Arg
         60                  65                  70

ACG AGT CGC CCT GGA AGG TGG CCA TGG CAG TGT TCT CTG CAG AGT GAA
290
Thr Ser Arg Pro Gly Arg Trp Pro Trp Gln Cys Ser Leu Gln Ser Glu
     75                  80                  85

CCC AGT GGA CAT ATC TGT GGC TGT GTC CTC ATT GCC AAG AAG TGG GTT
338
Pro Ser Gly His Ile Cys Gly Cys Val Leu Ile Ala Lys Lys Trp Val
 90                  95                  100                 105

GTG ACA GTT GCC CAC TGC TTC GAG GGG AGA GAG AAT GCT GCA GTT TGG
386
Val Thr Val Ala His Cys Phe Glu Gly Arg Glu Asn Ala Ala Val Trp
                110                 115                 120

AAA GTG GTG CTT GGC ATC AAC AAT CTA GAC CAT CCA TCA GTG TTC ATG
434
Lys Val Val Leu Gly Ile Asn Asn Leu Asp His Pro Ser Val Phe Met
            125                 130                 135

CAG ACA CGC TTT GTG AGG ACC ATC ATC CTG CAT CCC CGC TAC AGT CGA
482
```

```
Gln Thr Arg Phe Val Arg Thr Ile Ile Leu His Pro Arg Tyr Ser Arg
        140                 145                 150

GCA GTG GTG GAC TAT GAC ATC AGC ATC GTT GAG CTG AGT GAA GAC ATC
530
Ala Val Val Asp Tyr Asp Ile Ser Ile Val Glu Leu Ser Glu Asp Ile
    155                 160                 165

AGT GAG ACT GGC TAC GTC CGG CCT GTC TGC TTG CCC AAC CCG GAG CAG
578
Ser Glu Thr Gly Tyr Val Arg Pro Val Cys Leu Pro Asn Pro Glu Gln
170                 175                 180                 185

TGG CTA GAG CCT GAC ACG TAC TGC TAT ATC ACA GGC TGG GGC CAC ATG
626
Trp Leu Glu Pro Asp Thr Tyr Cys Tyr Ile Thr Gly Trp Gly His Met
                190                 195                 200

GGC AAT AAA ATG CCA TTT AAG CTG CAA GAG GGA GAG GTC CGC ATT ATT
674
Gly Asn Lys Met Pro Phe Lys Leu Gln Glu Gly Glu Val Arg Ile Ile
            205                 210                 215

TCT CTG GAA CAT TGT CAG TCC TAC TTT GAC ATG AAG ACC ATC ACC ACT
722
Ser Leu Glu His Cys Gln Ser Tyr Phe Asp Met Lys Thr Ile Thr Thr
        220                 225                 230

CGG ATG ATA TGT GCT GGC TAT GAG TCT GGC ACA GTT GAT TCA TGC ATG
770
Arg Met Ile Cys Ala Gly Tyr Glu Ser Gly Thr Val Asp Ser Cys Met
    235                 240                 245

GGT GAC TGG GGC GGT CCG TTG AAT TCT GT
799
Gly Asp Trp Gly Gly Pro Leu Asn Ser
250                 255
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 258 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Glu Pro Ser Val Thr Lys Leu Ile Gln Glu Gln Glu Lys Glu Pro Arg
  1               5                  10                  15

Trp Leu Thr Leu His Ser Asn Trp Glu Ser Leu Asn Gly Thr Thr Leu
             20                  25                  30

His Glu Leu Val Val Asn Gly Gln Ser Cys Glu Ser Arg Ser Lys Ile
         35                  40                  45

Ser Leu Leu Cys Thr Lys Gln Asp Cys Gly Arg Arg Pro Ala Ala Arg
     50                  55                  60

Met Asn Lys Arg Ile Leu Gly Gly Arg Thr Ser Arg Pro Gly Arg Trp
 65                  70                  75                  80

Pro Trp Gln Cys Ser Leu Gln Ser Glu Pro Ser Gly His Ile Cys Gly
                 85                  90                  95

Cys Val Leu Ile Ala Lys Lys Trp Val Val Thr Val Ala His Cys Phe
            100                 105                 110

Glu Gly Arg Glu Asn Ala Ala Val Trp Lys Val Val Leu Gly Ile Asn
        115                 120                 125

Asn Leu Asp His Pro Ser Val Phe Met Gln Thr Arg Phe Val Arg Thr
    130                 135                 140

Ile Ile Leu His Pro Arg Tyr Ser Arg Ala Val Val Asp Tyr Asp Ile
145                 150                 155                 160

Ser Ile Val Glu Leu Ser Glu Asp Ile Ser Glu Thr Gly Tyr Val Arg
                165                 170                 175

Pro Val Cys Leu Pro Asn Pro Glu Gln Trp Leu Glu Pro Asp Thr Tyr
            180                 185                 190

Cys Tyr Ile Thr Gly Trp Gly His Met Gly Asn Lys Met Pro Phe Lys
        195                 200                 205

Leu Gln Glu Gly Glu Val Arg Ile Ile Ser Leu Glu His Cys Gln Ser
    210                 215                 220
```

Tyr Phe Asp Met Lys Thr Ile Thr Thr Arg Met Ile Cys Ala Gly Tyr
225 230 235 240

Glu Ser Gly Thr Val Asp Ser Cys Met Gly Asp Trp Gly Gly Pro Leu
245 250 255

Asn Ser

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2241 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 166..1773

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATTTAATACG ACTCACTATA GGGAATTTGG CCCTCGAGGA AGAATTCGGC ACGAGGCTGC
60

GGCGCACTGT GAGGGAGTCG CTGTGATCCG GGGCCCCGAA CCCGACTGGA GCTGAAGCGC
120

AGGCTGCGGG GCGCGGAGTC GGGAGGCCTG AGTGTTCCTT CCAGC ATG TCG GAG
174
Met Ser Glu
1

GGG GAG TCC CAG ACA GTA CTT AGC AGT GGC TCA GAC CCA AAG GTA GAA
222
Gly Glu Ser Gln Thr Val Leu Ser Ser Gly Ser Asp Pro Lys Val Glu
5 10 15

TCT TCA TCT TCA GCT CCT GGC CTG ACA TCA GTG TCA CCT CCT GTG ACC
270

```
Ser Ser Ser Ser Ala Pro Gly Leu Thr Ser Val Ser Pro Pro Val Thr
 20                  25                  30                  35

TCC ACA ACC TCA GCT GCT TCC CCA GAG GAA GAA GAA GAA AGT GAA GAT
318
Ser Thr Thr Ser Ala Ala Ser Pro Glu Glu Glu Glu Ser Glu Asp
                 40                  45                  50

GAG TCT GAG ATT TTG GAA GAG TCG CCC TGT GGG CGC TGG CAG AAG AGG
366
Glu Ser Glu Ile Leu Glu Glu Ser Pro Cys Gly Arg Trp Gln Lys Arg
             55                  60                  65

CGA GAA GAG GTG AAT CAA CGG AAT GTA CCA GGT ATT GAC AGT GCA TAC
414
Arg Glu Glu Val Asn Gln Arg Asn Val Pro Gly Ile Asp Ser Ala Tyr
         70                  75                  80

CTG GCC ATG GAT ACA GAG GAA GGT GTA GAG GTT GTG TGG AAT GAG GTA
462
Leu Ala Met Asp Thr Glu Glu Gly Val Glu Val Val Trp Asn Glu Val
     85                  90                  95

CAG TTC TCT GAA CGC AAG AAC TAC AAG CTG CAG GAG GAA AAG GTT TGT
510
Gln Phe Ser Glu Arg Lys Asn Tyr Lys Leu Gln Glu Glu Lys Val Cys
100                 105                 110                 115

GCT GTG TTT GAT AAT TTG ATT CAA TTG GAG CAT CTT AAC ATT GTT AAG
558
Ala Val Phe Asp Asn Leu Ile Gln Leu Glu His Leu Asn Ile Val Lys
                120                 125                 130

TTT CAC AAA TAT TGG GCT GAC ATT AAA GAG AAC AAG GCC AGG GTC ATT
606
Phe His Lys Tyr Trp Ala Asp Ile Lys Glu Asn Lys Ala Arg Val Ile
            135                 140                 145

TTT ATC ACA GGA TAC ATG TCA TCT GGG AGT CTG AAG CAA TTT CTG AAG
654
Phe Ile Thr Gly Tyr Met Ser Ser Gly Ser Leu Lys Gln Phe Leu Lys
        150                 155                 160

AAG ACC CAA AAG AAC CAC CAG ACG ATG AAT GAA AAG GCA TGG AAG CGT
702
```

```
Lys Thr Gln Lys Asn His Gln Thr Met Asn Glu Lys Ala Trp Lys Arg
    165                 170                 175

TGG TGC ACA CAA ATC CTC TCT GCC CTA AGC TAC CTG CAC TCC TGT GAC
750
Trp Cys Thr Gln Ile Leu Ser Ala Leu Ser Tyr Leu His Ser Cys Asp
180                 185                 190                 195

CCC CCC ATC ATC CAT GGG AAC CTG ACC TGT GAC ACC ATC TTC ATC CAG
798
Pro Pro Ile Ile His Gly Asn Leu Thr Cys Asp Thr Ile Phe Ile Gln
                200                 205                 210

CAC AAC GGA CTC ATC AAG ATT GGC TCT GTG GCT CCT GAC ACT ATC AAC
846
His Asn Gly Leu Ile Lys Ile Gly Ser Val Ala Pro Asp Thr Ile Asn
            215                 220                 225

AAT CAT GTG AAG ACT TGT CGA GAA GAG CAG AAG AAT CTA CAC TTC TTT
894
Asn His Val Lys Thr Cys Arg Glu Glu Gln Lys Asn Leu His Phe Phe
        230                 235                 240

GCA CCA GAG TAT GGA GAA GTC ACT AAT GTG ACA ACA GCA GTG GAC ATC
942
Ala Pro Glu Tyr Gly Glu Val Thr Asn Val Thr Thr Ala Val Asp Ile
    245                 250                 255

TAC TCC TTT GGC ATG TGT GCA CTG GGG ATG GCA GTG CTG GAG ATT CAG
990
Tyr Ser Phe Gly Met Cys Ala Leu Gly Met Ala Val Leu Glu Ile Gln
260                 265                 270                 275

GGC AAT GGA GAG TCC TCA TAT GTG CCA CAG GAA GCC ATC AGC AGT GCC
1038
Gly Asn Gly Glu Ser Ser Tyr Val Pro Gln Glu Ala Ile Ser Ser Ala
                280                 285                 290

ATC CAG CTT CTA GAA GAC CCA TTA CAG AGG GAG TTC ATT CAA AAG TGC
1086
Ile Gln Leu Leu Glu Asp Pro Leu Gln Arg Glu Phe Ile Gln Lys Cys
            295                 300                 305

CTG CAG TCT GAG CCT GCT CGC AGA CCA ACA GCC AGA GAA CTT CTG TTC
1134
```

```
Leu Gln Ser Glu Pro Ala Arg Arg Pro Thr Ala Arg Glu Leu Leu Phe
        310                 315                 320

CAC CCA GCA TTG TTT GAA GTG CCC TCG CTC AAA CTC CTT GCG GCC CAC
1182
His Pro Ala Leu Phe Glu Val Pro Ser Leu Lys Leu Leu Ala Ala His
    325                 330                 335

TGC ATT GTG GGA CAC CAA CAC ATG ATC CCA GAG AAC GCT CTA GAG GAG
1230
Cys Ile Val Gly His Gln His Met Ile Pro Glu Asn Ala Leu Glu Glu
340                 345                 350                 355

ATC ACC AAA AAC ATG GAT ACT AGT GCC GTA CTG GCT GAA ATC CCT GCA
1278
Ile Thr Lys Asn Met Asp Thr Ser Ala Val Leu Ala Glu Ile Pro Ala
                360                 365                 370

GGA CCA GGA AGA GAA CCA GTT CAG ACT TTG TAC TCT CAG TCA CCA GCT
1326
Gly Pro Gly Arg Glu Pro Val Gln Thr Leu Tyr Ser Gln Ser Pro Ala
            375                 380                 385

CTG GAA TTA GAT AAA TTC CTT GAA GAT GTC AGG AAT GGG ATC TAT CCT
1374
Leu Glu Leu Asp Lys Phe Leu Glu Asp Val Arg Asn Gly Ile Tyr Pro
        390                 395                 400

CTG ACA GCC TTT GGG CTG CCT CGG CCC CAG CAG CCA CAG CAG GAG GAG
1422
Leu Thr Ala Phe Gly Leu Pro Arg Pro Gln Gln Pro Gln Gln Glu Glu
    405                 410                 415

GTG ACA TCA CCT GTC GTG CCC CCC TCT GTC AAG ACT CCG ACA CCT GAA
1470
Val Thr Ser Pro Val Val Pro Pro Ser Val Lys Thr Pro Thr Pro Glu
420                 425                 430                 435

CCA GCT GAG GTG GAG ACT CGC AAG GTG GTG CTG ATG CAG TGC AAC ATT
1518
Pro Ala Glu Val Glu Thr Arg Lys Val Val Leu Met Gln Cys Asn Ile
                440                 445                 450

GAG TCG GTG GAG GAG GGA GTC AAA CAC CAC CTG ACA CTT CTG CTG AAG
1566
```

```
Glu Ser Val Glu Glu Gly Val Lys His His Leu Thr Leu Leu Leu Lys
            455                 460                 465

TTG GAG GAC AAA CTG AAC CGG CAC CTG AGC TGT GAC CTG ATG CCA AAT
1614
Leu Glu Asp Lys Leu Asn Arg His Leu Ser Cys Asp Leu Met Pro Asn
        470                 475                 480

GAG AAT ATC CCC GAG TTG GCG GCT GAG CTG GTG CAG CTG GGC TTC ATT
1662
Glu Asn Ile Pro Glu Leu Ala Ala Glu Leu Val Gln Leu Gly Phe Ile
    485                 490                 495

AGT GAG GCT GAC CAG AGC CGG TTG ACT TCT CTG CTA GAA GAG ACC TTG
1710
Ser Glu Ala Asp Gln Ser Arg Leu Thr Ser Leu Leu Glu Glu Thr Leu
500                 505                 510                 515

AAC AAG TTC AAT TTT GCC AGG AAC AGT ACC CTC AAC TCA GCC GCT GTC
1758
Asn Lys Phe Asn Phe Ala Arg Asn Ser Thr Leu Asn Ser Ala Ala Val
                520                 525                 530

ACC GTC TCC TCT TAGAGCTCAC TCGGGCCAGG CCCTGATCTG CGCTGTGGCT
1810
Thr Val Ser Ser
            535

GTCCCTGGAC GTGCTGCAGC CCTCCTGTCC CTTCCCCCCA GTCAGTATTA CCCTGTGAAG
1870

CCCCTTCCCT CCTTTATTAT TCAGGAGGGC TGGGGGGGCT CCCTGGTTCT GAGCATCATC
1930

CTTTCCCCTC CCCTCTCTTC CTCCCCTCTG CACTTTGTTT ACTTGTTTTG CACAGACGTG
1990

GGCCTGGGCC TTCTCAGCAG CCGCCTTCTA GTTGGGGGCT AGTCGCTGAT CTGCCGGCTC
2050

CCGCCCAGCC TGTGTGGAAA GGAGGCCCAC GGGCACTAGG GGAGCCGAAT TCTACAATCC
2110

CGCTGGGGCG GCCGGGGCGG GAGAGAAAGG TGGTGCTGCA GTGGTGGCCC TGGGGGGCCA
2170
```

TTCGATTCGC CTCAGTTGCT GCTGTAATAA AAGTCTACTT TTTGCTAAAA AAAAAAAAAA
2230

AAAAAAAAAA A
2241

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 535 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Ser Glu Gly Glu Ser Gln Thr Val Leu Ser Ser Gly Ser Asp Pro
1 5 10 15

Lys Val Glu Ser Ser Ser Ser Ala Pro Gly Leu Thr Ser Val Ser Pro
20 25 30

Pro Val Thr Ser Thr Thr Ser Ala Ala Ser Pro Glu Glu Glu Glu Glu
35 40 45

Ser Glu Asp Glu Ser Glu Ile Leu Glu Glu Ser Pro Cys Gly Arg Trp
50 55 60

Gln Lys Arg Arg Glu Glu Val Asn Gln Arg Asn Val Pro Gly Ile Asp
65 70 75 80

Ser Ala Tyr Leu Ala Met Asp Thr Glu Glu Gly Val Glu Val Val Trp
85 90 95

Asn Glu Val Gln Phe Ser Glu Arg Lys Asn Tyr Lys Leu Gln Glu Glu
100 105 110

Lys Val Cys Ala Val Phe Asp Asn Leu Ile Gln Leu Glu His Leu Asn
115 120 125

Ile Val Lys Phe His Lys Tyr Trp Ala Asp Ile Lys Glu Asn Lys Ala
130 135 140

```
Arg Val Ile Phe Ile Thr Gly Tyr Met Ser Ser Gly Ser Leu Lys Gln
145                 150                 155                 160

Phe Leu Lys Lys Thr Gln Lys Asn His Gln Thr Met Asn Glu Lys Ala
                165                 170                 175

Trp Lys Arg Trp Cys Thr Gln Ile Leu Ser Ala Leu Ser Tyr Leu His
            180                 185                 190

Ser Cys Asp Pro Pro Ile Ile His Gly Asn Leu Thr Cys Asp Thr Ile
        195                 200                 205

Phe Ile Gln His Asn Gly Leu Ile Lys Ile Gly Ser Val Ala Pro Asp
    210                 215                 220

Thr Ile Asn Asn His Val Lys Thr Cys Arg Glu Glu Gln Lys Asn Leu
225                 230                 235                 240

His Phe Phe Ala Pro Glu Tyr Gly Glu Val Thr Asn Val Thr Thr Ala
                245                 250                 255

Val Asp Ile Tyr Ser Phe Gly Met Cys Ala Leu Gly Met Ala Val Leu
            260                 265                 270

Glu Ile Gln Gly Asn Gly Glu Ser Ser Tyr Val Pro Gln Glu Ala Ile
        275                 280                 285

Ser Ser Ala Ile Gln Leu Leu Glu Asp Pro Leu Gln Arg Glu Phe Ile
    290                 295                 300

Gln Lys Cys Leu Gln Ser Glu Pro Ala Arg Arg Pro Thr Ala Arg Glu
305                 310                 315                 320

Leu Leu Phe His Pro Ala Leu Phe Glu Val Pro Ser Leu Lys Leu Leu
                325                 330                 335

Ala Ala His Cys Ile Val Gly His Gln His Met Ile Pro Glu Asn Ala
            340                 345                 350

Leu Glu Glu Ile Thr Lys Asn Met Asp Thr Ser Ala Val Leu Ala Glu
        355                 360                 365

Ile Pro Ala Gly Pro Gly Arg Glu Pro Val Gln Thr Leu Tyr Ser Gln
    370                 375                 380
```

```
Ser Pro Ala Leu Glu Leu Asp Lys Phe Leu Glu Asp Val Arg Asn Gly
385                 390                 395                 400

Ile Tyr Pro Leu Thr Ala Phe Gly Leu Pro Arg Pro Gln Gln Pro Gln
                405                 410                 415

Gln Glu Glu Val Thr Ser Pro Val Val Pro Pro Ser Val Lys Thr Pro
            420                 425                 430

Thr Pro Glu Pro Ala Glu Val Glu Thr Arg Lys Val Val Leu Met Gln
        435                 440                 445

Cys Asn Ile Glu Ser Val Glu Glu Gly Val Lys His His Leu Thr Leu
    450                 455                 460

Leu Leu Lys Leu Glu Asp Lys Leu Asn Arg His Leu Ser Cys Asp Leu
465                 470                 475                 480

Met Pro Asn Glu Asn Ile Pro Glu Leu Ala Ala Glu Leu Val Gln Leu
                485                 490                 495

Gly Phe Ile Ser Glu Ala Asp Gln Ser Arg Leu Thr Ser Leu Leu Glu
            500                 505                 510

Glu Thr Leu Asn Lys Phe Asn Phe Ala Arg Asn Ser Thr Leu Asn Ser
        515                 520                 525

Ala Ala Val Thr Val Ser Ser
    530                 535
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCACAGTCGA CCAAGCCGGA GTCGCAGAG 39

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCACAAAGCT TGCCAGGAGG GGTCTGGCTG 30

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCACAACCAT GGCCAAGCCG GAGTCGCAGG AG 32

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCACAAGATC TCCAGGAGGG GTCTGGCTG 29

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Lys Pro Glu Ser Gln Glu Ala Ala Pro Leu Ser Gly Pro Cys
5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Glu Asp Ala Glu Leu Gly Arg Trp Pro Trp Gln Gly Ser Leu Arg Leu Trp Asp
5 10 15
Cys

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Gly Tyr Ile Lys Glu Asp Glu Ala Leu Pro Ser Pro His Thr Leu Gln Cys
5 10 15

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCACAGGTAC CGAGGCCATG GGCGCGCGC 29

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GCACATCTAG ATCAGTGGTG GTGGTGGTGG TGGACCGGCC CCAGGAGTGG 50

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCACAGCGGC CGCGAGGCCA TGGGCGCGCG C 31

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GCACAGCGGC CGCTCAGTGG TGGTGGTGGT GGTGCCAGGA GGGGTCTGGC TG 52

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTGACTTCCA TGCCATCCTT 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GCTCACGACT CCAATCTGAT 20

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Arg Ile Val Gly Gly
5

DATED this 18th day of November, 1997

Amrad Operations Pty Ltd

~~The Council of The Queensland Institute of Medical Research~~

By DAVIES COLLISON CAVE

Patent Attorneys for the Applicants

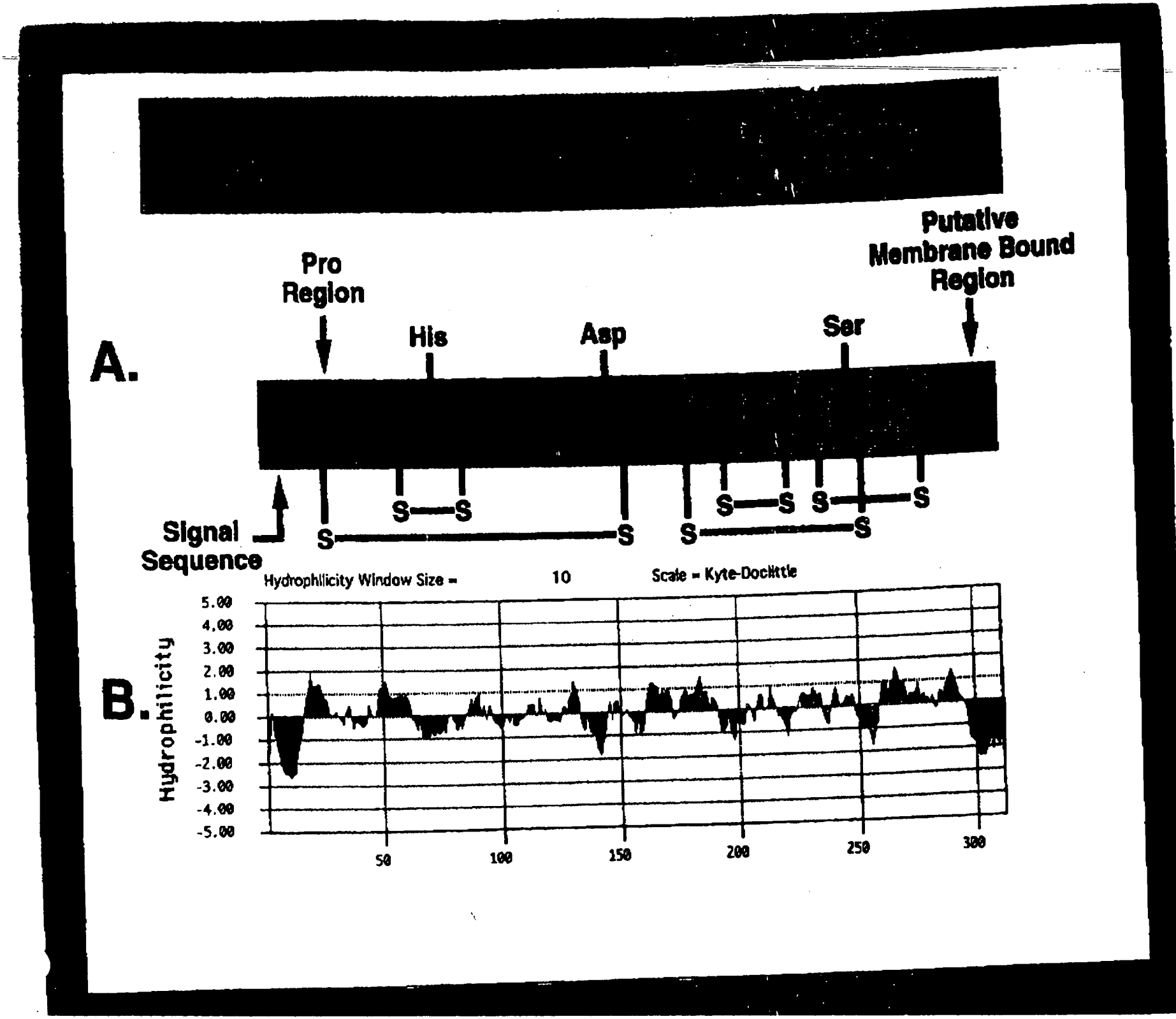

A.
Pro Region
His
Asp
Ser
Putative Membrane Bound Region
Signal Sequence
S
B.
Hydrophilicity Window Size = 10
Scale = Kyte-Doolittle
Hydrophilicity
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
-4.00
-5.00
50
100
150
200
250
300

FIGURE 1

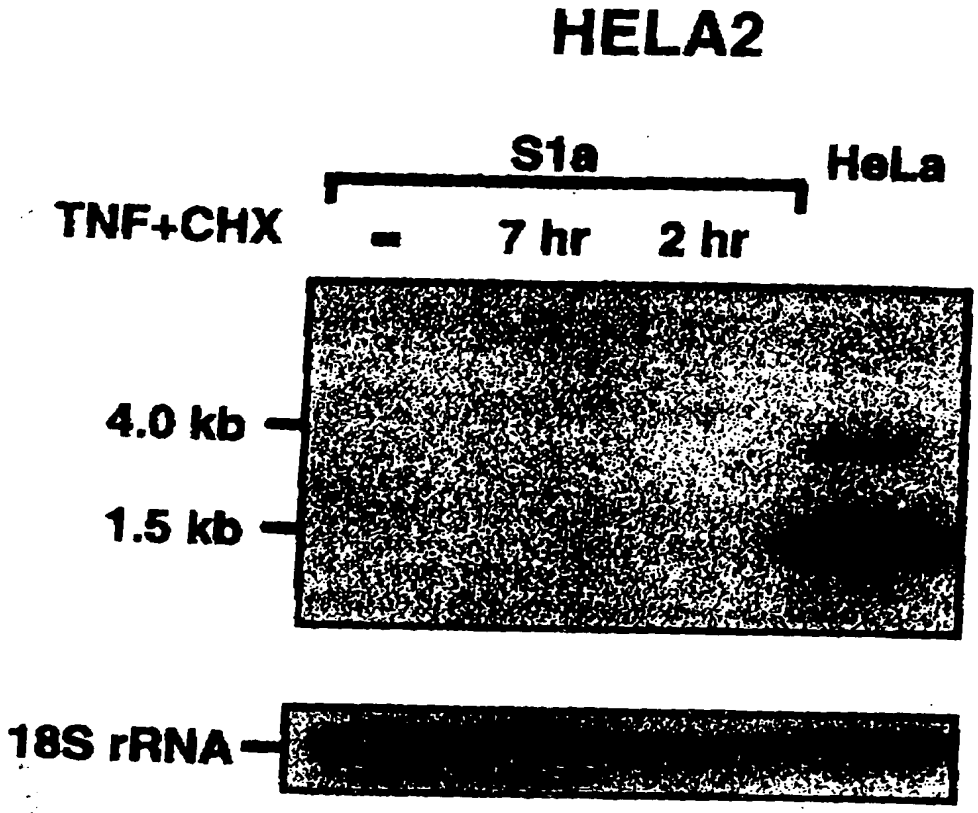


FIGURE 2

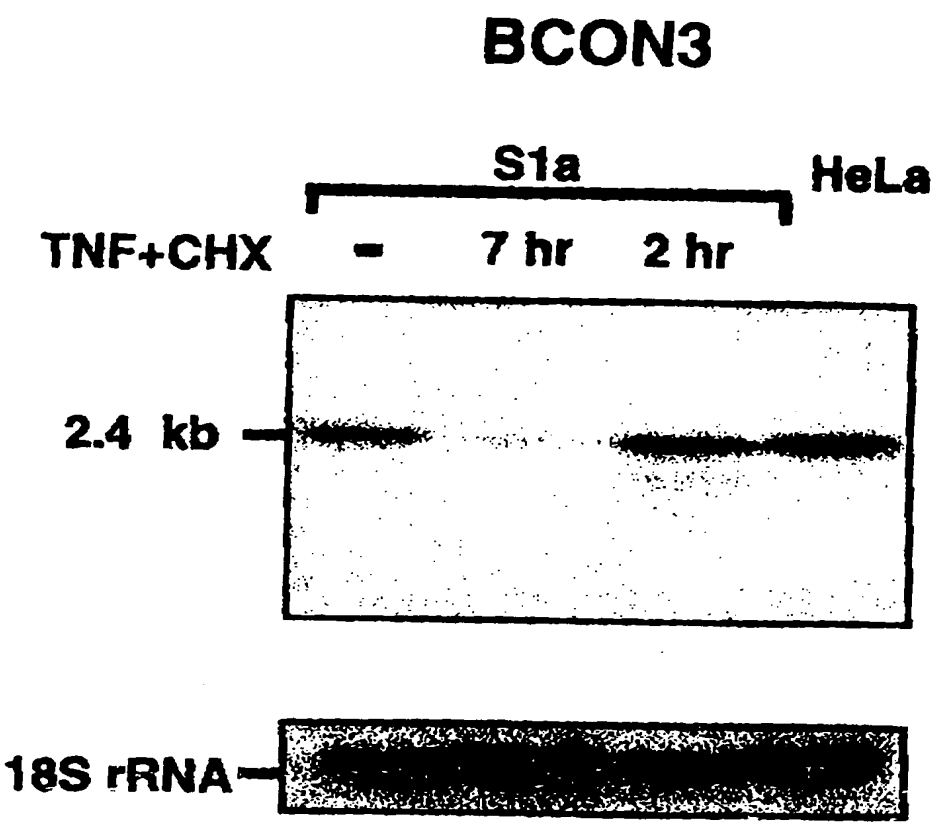
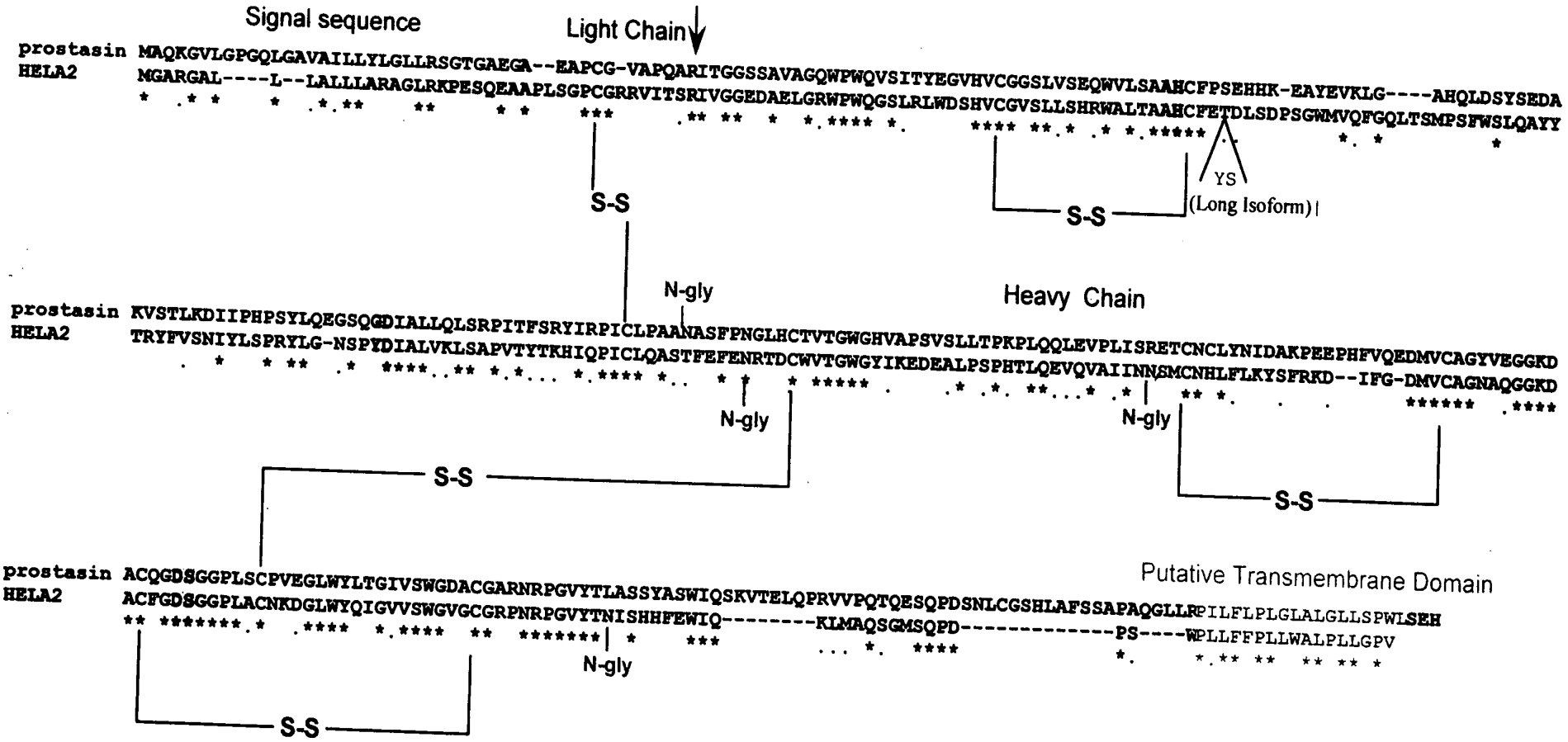


FIGURE 3


Signal sequence
Light Chain
prostasin MAQKGVLGPGQLGAVAILLYLGLLRSGTGAEGA--EAPCG-VAPQARITGGSSAVAGQWPWQVSITYEGVHVCGGSLVSEQWVLSAAHCFPSEHHK-EAYEVKLG----AHQLDSYSEDA
HELA2 MGARGAL----L--LALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETDLSDPSGWMVQFGQLTSMPSFWSLQAYY
S-S
S-S
YS
(Long Isoform)
N-gly
Heavy Chain
prostasin KVSTLKDIIPHPSYLQEGSQGDIALLQLSRPITFSRYIRPICLPAANASFPNGLHCTVTGWGHVAPSVSLLTPKPLQQLEVPLISRETCNCLYNIDAKPEEPHFVQEDMVCAGYVEGGKD
HELA2 TRYFVSNIYLSPRYLG-NSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKD--IFG-DMVCAGNAQGGKD
N-gly
N-gly
S-S
S-S
prostasin ACQGDSGGPLSCPVEGLWYLTGIVSWGDACGARNRPGVYTLASSYASWIQSKVTELQPRVVPQTQESQPDSNLCGSHLAFSSAPAQGLLRPILFLPLGLALGLLSPWLSEH
HELA2 ACFGDSGGPLACNKDGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEWIQ--------KLMAQSGMSQPD----------PS----WPLLFFPLLWALPLLGPV
Putative Transmembrane Domain
N-gly
S-S

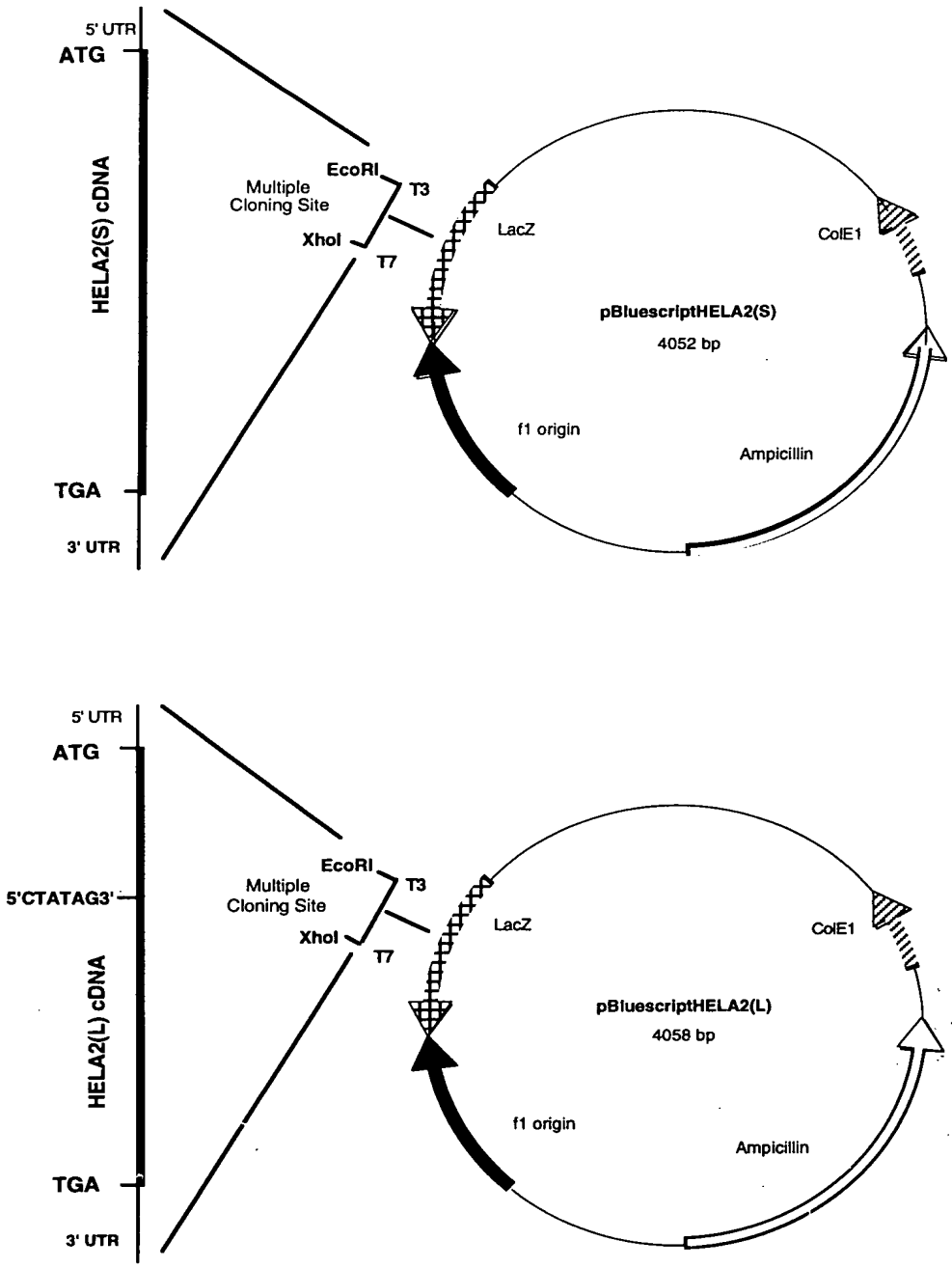

5' UTR
ATG
HELA2(S) cDNA
TGA
3' UTR
EcoRI
T3
Multiple
Cloning Site
XhoI
T7
LacZ
ColE1
pBluescriptHELA2(S)
4052 bp
f1 origin
Ampicillin
5' UTR
ATG
5'CTATAG3'
HELA2(L) cDNA
TGA
3' UTR
EcoRI
T3
Multiple
Cloning Site
XhoI
T7
LacZ
ColE1
pBluescriptHELA2(L)
4058 bp
f1 origin
Ampicillin

FIGURE 5

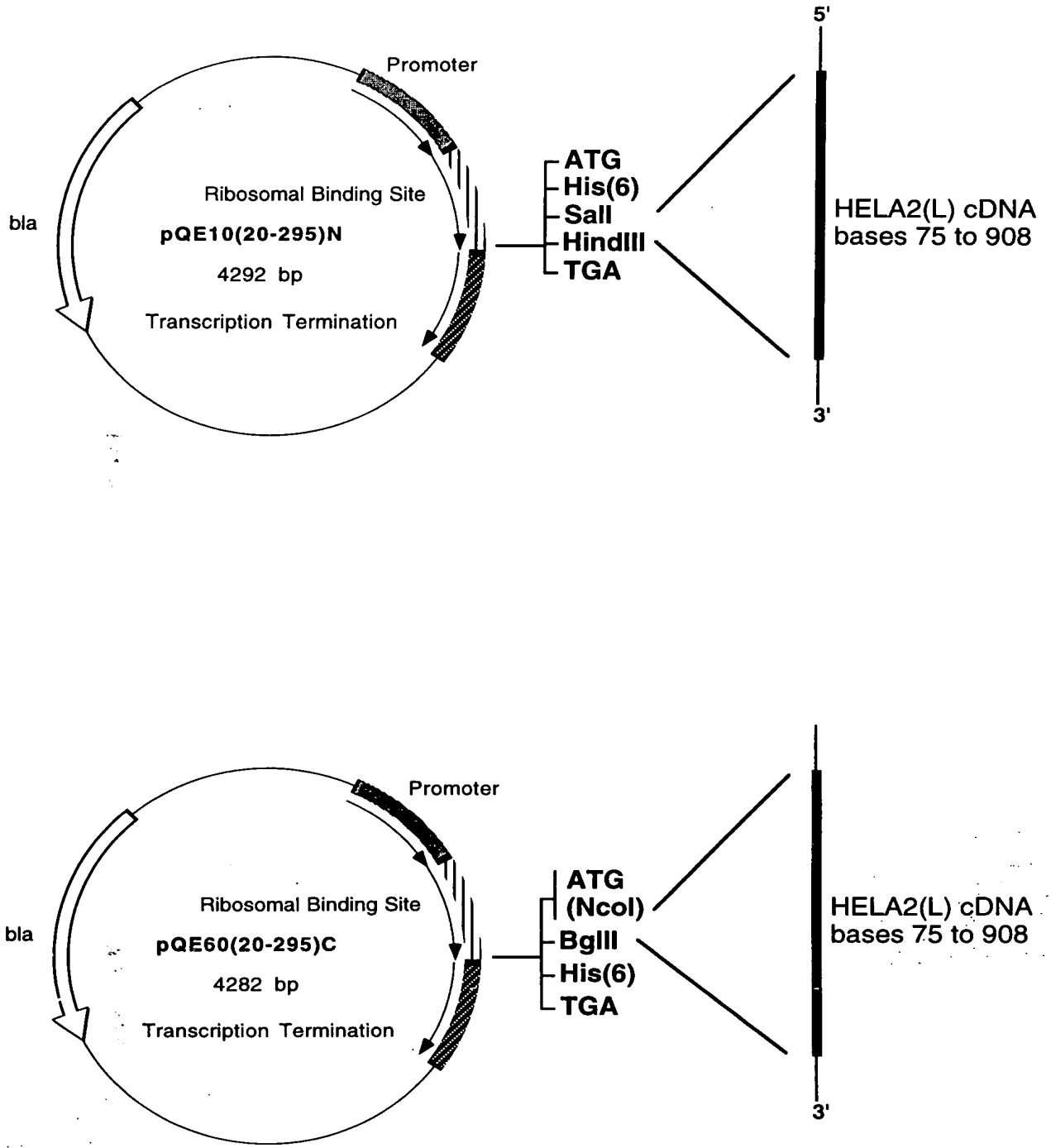

Promoter
Ribosomal Binding Site
bla
pQE10(20-295)N
4292 bp
Transcription Termination
ATG
His(6)
SalI
HindIII
TGA
5'
3'
HELA2(L) cDNA
bases 75 to 908
Promoter
Ribosomal Binding Site
bla
pQE60(20-295)C
4282 bp
Transcription Termination
ATG
(NcoI)
BglII
His(6)
TGA
5'
3'
HELA2(L) cDNA
bases 75 to 908

FIGURE 6

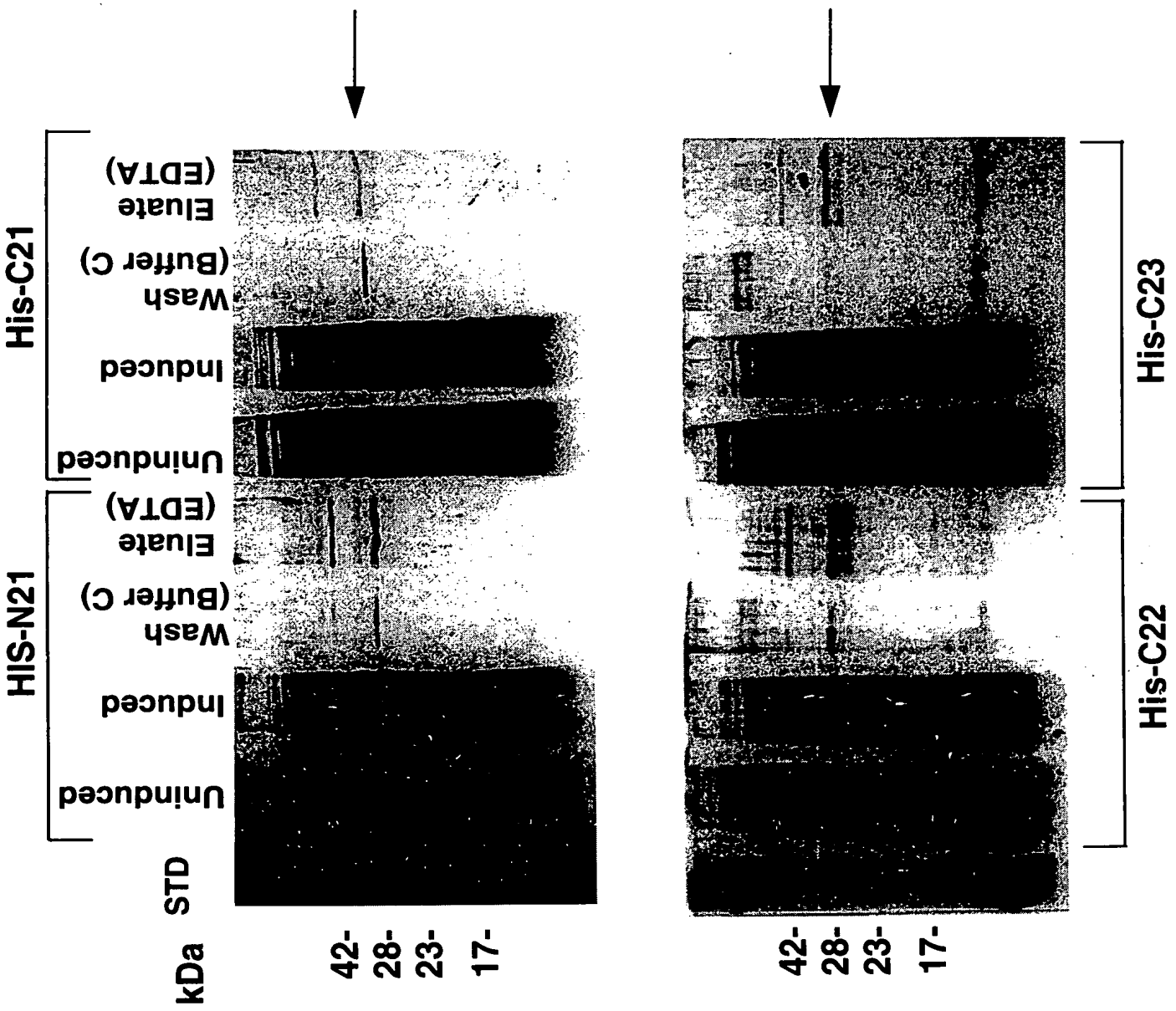

kDa
STD
42-
28-
23-
17-
Uninduced
Induced
Wash (Buffer C)
Eluate (EDTA)
HIS-N21
His-C21
His-C22
His-C23

FIGURE 7A

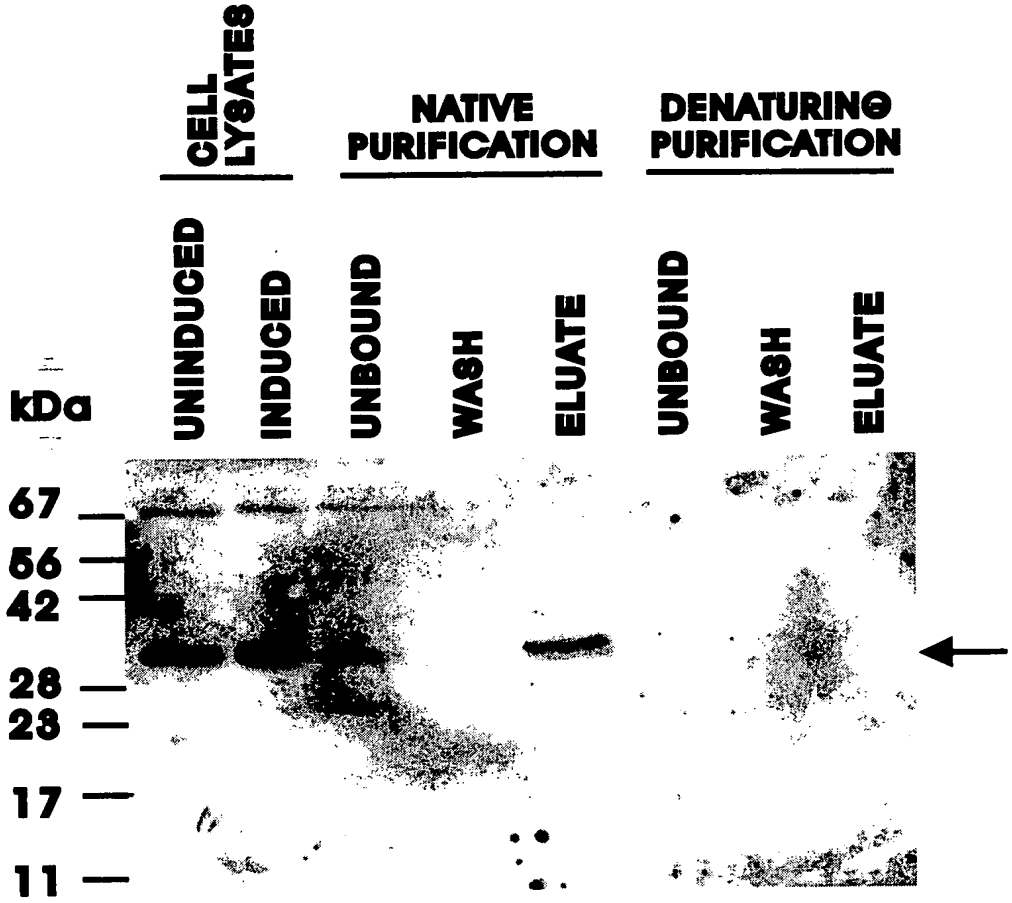

CELL LYSATES
NATIVE PURIFICATION
DENATURING PURIFICATION
UNINDUCED
INDUCED
UNBOUND
WASH
ELUATE
UNBOUND
WASH
ELUATE
kDa
67
56
42
28
28
17
11

FIGURE 7B

```
1                                                        GCCGCGGGAGAGGAGGCC
19   ATGGGCGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCTGGACTCAGGAAG
      M  G  A  R  G  A  L  L  L  A  L  L  L  A  R  A  G  L  R  K    20

79   CCGGAGTCGCAGGAGGCGGCGCCGTTATCAGGACCATGCGGCCGACGGGTCATCACGTCG
      P  E  S  Q  E  A  A  P  L  S  G  P  C  G  R  R  V  I  T  S    40

139  CGCATCGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGC
      R  I  V  G  G  E  D  A  E  L  G  R  W  P  W  Q  G  S  L  R    60

199  CTGTGGGATTCCCACGTATGCGGAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCG
      L  W  D  S  H  V  C  G  V  S  L  L  S  H  R  W  A  L  T  A    80

259  GCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCCTCCGGGTGGATGGTCCAGTTT
      A  H  C  F  E  T  Y  S  D  L  S  D  P  S  G  W  M  V  Q  F   100

319  GGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTACTTC
      G  Q  L  T  S  M  P  S  F  W  S  L  Q  A  Y  Y  T  R  Y  F   120

379  GTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTG
      V  S  N  I  Y  L  S  P  R  Y  L  G  N  S  P  Y  D  I  A  L   140

439  GTGAAGCTGTCTGCACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCC
      V  K  L  S  A  P  V  T  Y  T  K  H  I  Q  P  I  C  L  Q  A   160

499  TCCACATTTGAGTTTGAGAACCGGACAGACTGCTGGGTGACTGGCTGGGGGTACATCAAA
      S  T  F  E  F  E  N  R  T  D  C  W  V  T  G  W  G  Y  I  K   180

559  GAGGATGAGGCACTGCCATCTCCCCACACCCTCCAGGAAGTTCAGGTCGCCATCATAAAC
      E  D  E  A  L  P  S  P  H  T  L  Q  E  V  Q  V  A  I  I  N   200

619  AACTCTATGTGCAACCACCTCTTCCTCAAGTACAGTTTCCGCAAGGACATCTTTGGAGAC
      N  S  M  C  N  H  L  F  L  K  Y  S  F  R  K  D  I  F  G  D   220

679  ATGGTTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTTCGGTGACTCAGGTGGA
      M  V  C  A  G  N  A  Q  G  G  K  D  A  C  F  G  D  S  G  G   240

739  CCCTTGGCCTGTAACAAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTG
      P  L  A  C  N  K  N  G  L  W  Y  Q  I  G  V  V  S  W  G  V   260

799  GGCTGTGGTCGGCCCAATCGGCCCGGTGTCTACACCAATATCAGCCACCACTTTGAGTGG
      G  C  G  R  P  N  R  P  G  V  Y  T  N  I  S  H  H  F  E  W   280

859  ATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCCTCCTGGCCGCTACTC
      I  Q  K  L  M  A  Q  S  G  M  S  Q  P  D  P  S  W  P  L  L   300

919  TTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGAGCCTACCTGAGCCCA
      F  F  P  L  L  W  A  L  P  L  L  G  P  V  *                  314

979  TGCAGCCTGGGGCCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAA
1039 ACACATTCCAGTTGATGCCTTGCAGGGCATTCTTCAAAAAAAAAAAAAAAAAAAAAAAA
1099 AAAAAAAAAAAAAAAAAA
```

FIGURE 8

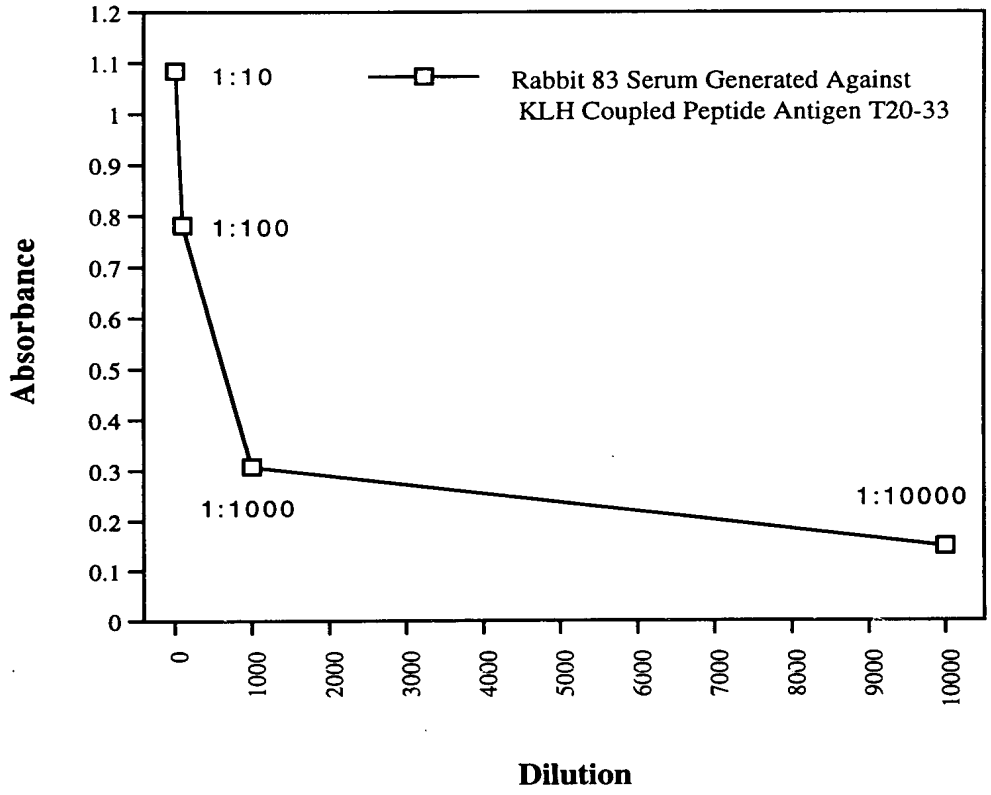

Absorbance
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1:10
1:100
1:1000
1:10000
Rabbit 83 Serum Generated Against
KLH Coupled Peptide Antigen T20-33
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Dilution

FIGURE 9

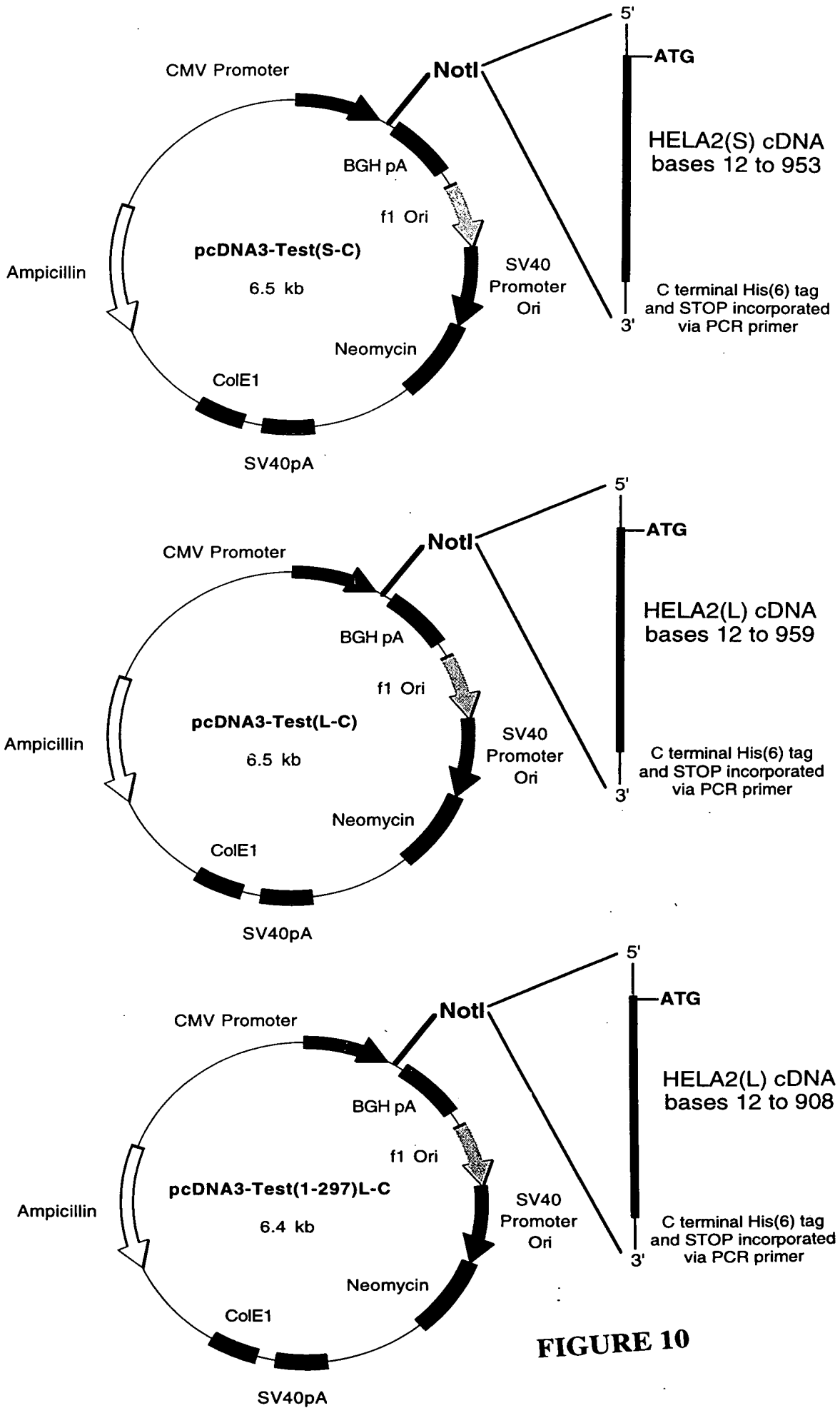

5'
ATG
CMV Promoter
NotI
HELA2(S) cDNA
bases 12 to 953
BGH pA
f1 Ori
pcDNA3-Test(S-C)
Ampicillin
6.5 kb
SV40
Promoter
Ori
C terminal His(6) tag
and STOP incorporated
via PCR primer
3'
Neomycin
ColE1
SV40pA
5'
ATG
CMV Promoter
NotI
HELA2(L) cDNA
bases 12 to 959
BGH pA
f1 Ori
pcDNA3-Test(L-C)
Ampicillin
6.5 kb
SV40
Promoter
Ori
C terminal His(6) tag
and STOP incorporated
via PCR primer
3'
Neomycin
ColE1
SV40pA
5'
ATG
CMV Promoter
NotI
HELA2(L) cDNA
bases 12 to 908
BGH pA
f1 Ori
pcDNA3-Test(1-297)L-C
Ampicillin
6.4 kb
SV40
Promoter
Ori
C terminal His(6) tag
and STOP incorporated
via PCR primer
3'
Neomycin
ColE1
SV40pA

FIGURE 10

TESTISIN (HELA2)

whole brain	amygdala	caudate nucl.	cerebellum	c. cortex	fr. lobe	hippocamp.	med. obl.
occ. lobes	putamen	subst.nigra.	temp. lobe	thalamus	sub.th. nuc.	spinal cord	
heart	aorta	skel.musc.	colon	bladder	uterus	prostate	stomach
testis	ovary	pancreas	pit. gland	adr. gland	thyr. gland	saliv. gland	mam.gland
kidney	liver	sm. intest.	spleen	thymus	periph. leuk.	lymph node	bone marr.
appendix	lung	trachea	placenta				
fetal brain	fetal heart	fetal kidney	fetal liver	fetal spleen	fetal thymus	fetal lung	
yeast RNA	yeast tRNA	E.coli RNA	E.coli DNA	Poly r(A)	hum. Cot DNA	hum DNA 100ng	hum DNA 500ng

FIGURE 11A

BCON3

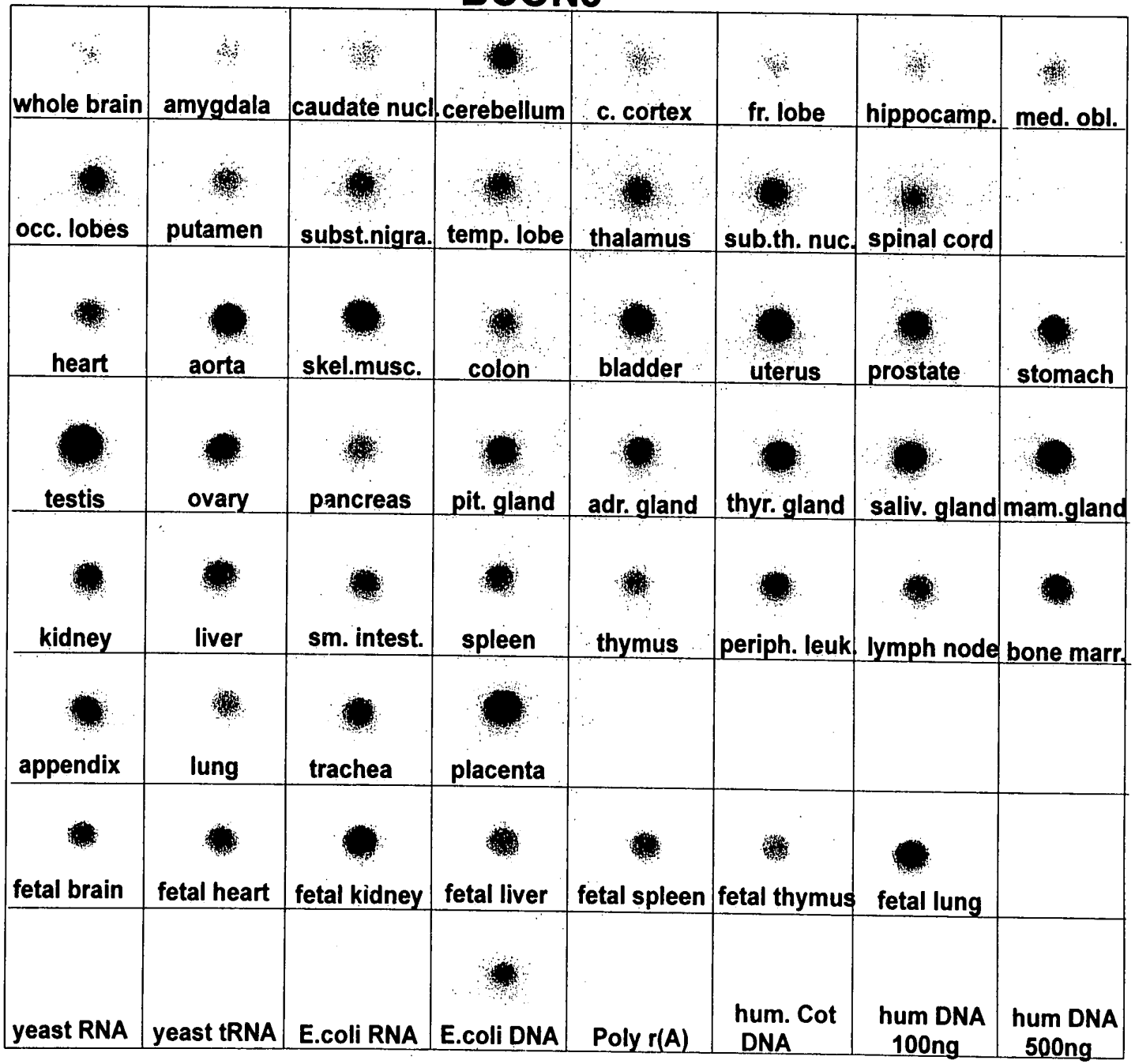
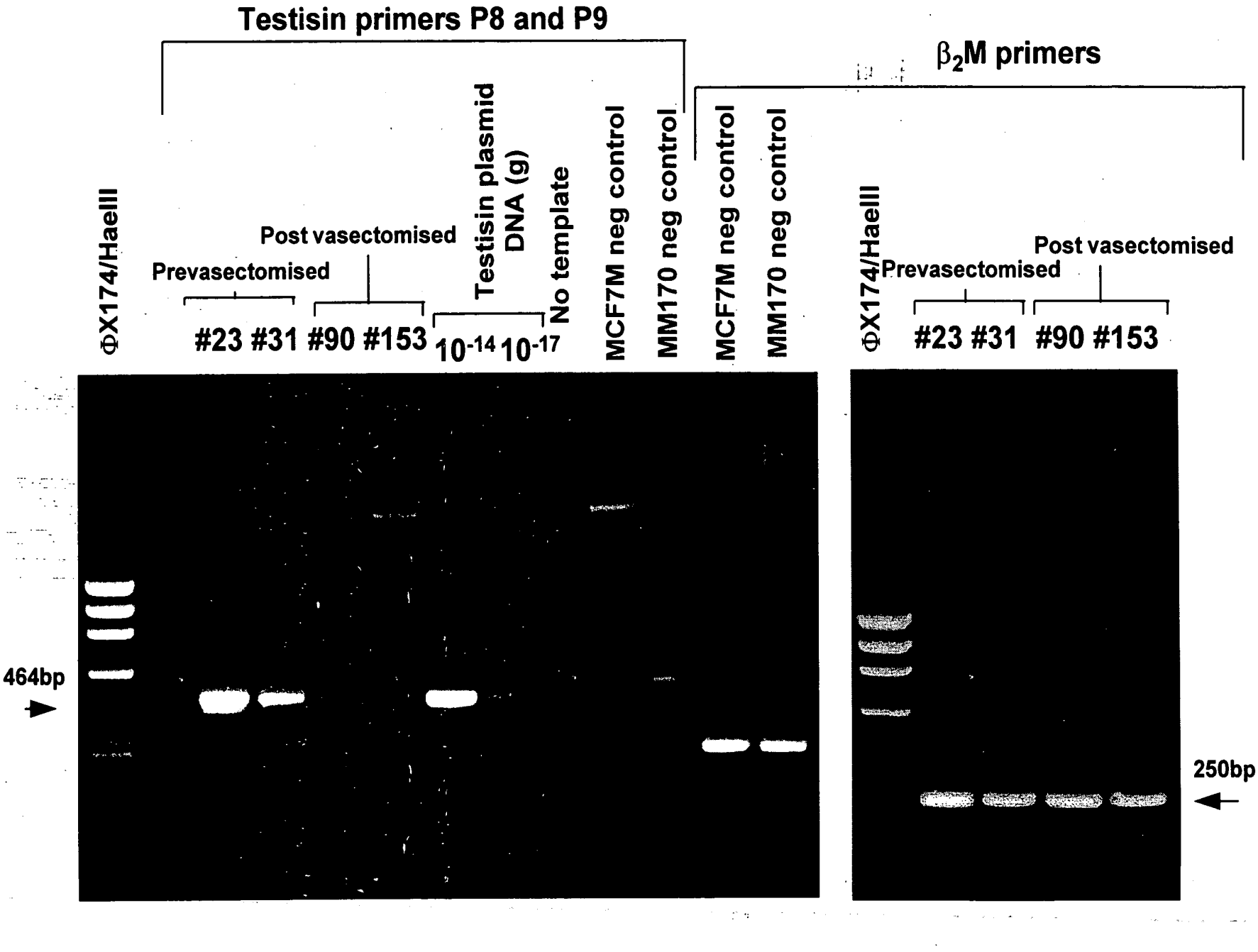


FIGURE 11B

FIGURE 12


Testisin primers P8 and P9
β2M primers
ΦX174/HaeIII
Prevasectomised
Post vasectomised
#23 #31 #90 #153
Testisin plasmid DNA (g)
10-14 10-17
No template
MCF7M neg control
MM170 neg control
MCF7M neg control
MM170 neg control
ΦX174/HaeIII
Prevasectomised
Post vasectomised
#23 #31 #90 #153
464bp
250bp

BY CLONTECH "MASTER" BLOT

(+) Testes

(+/-) Salivary Gland

BY NORTHERN BLOT

(+)	HeLa Cells	(cervical cancer)
(+)	253-3D Cells	(melanoma)
(+)	U937	(monoblast like)
(-)	other lines tested and negative:	SW480; SW620; MCF7; MCF7M; HaCat; MM6; 293; MM170; MM229; MM418c!; MM418c5; MM96L.

BY EST DATABASE MATCH

(+) Colon villous adenoma

(+) Pancreatic adenocarcinoma cell line

(+) Malignant prostate cancer cells

(+) Ovarian tumours

FIGURE 13

Testisin (HELA2) is located on human chromosome 16p13.3

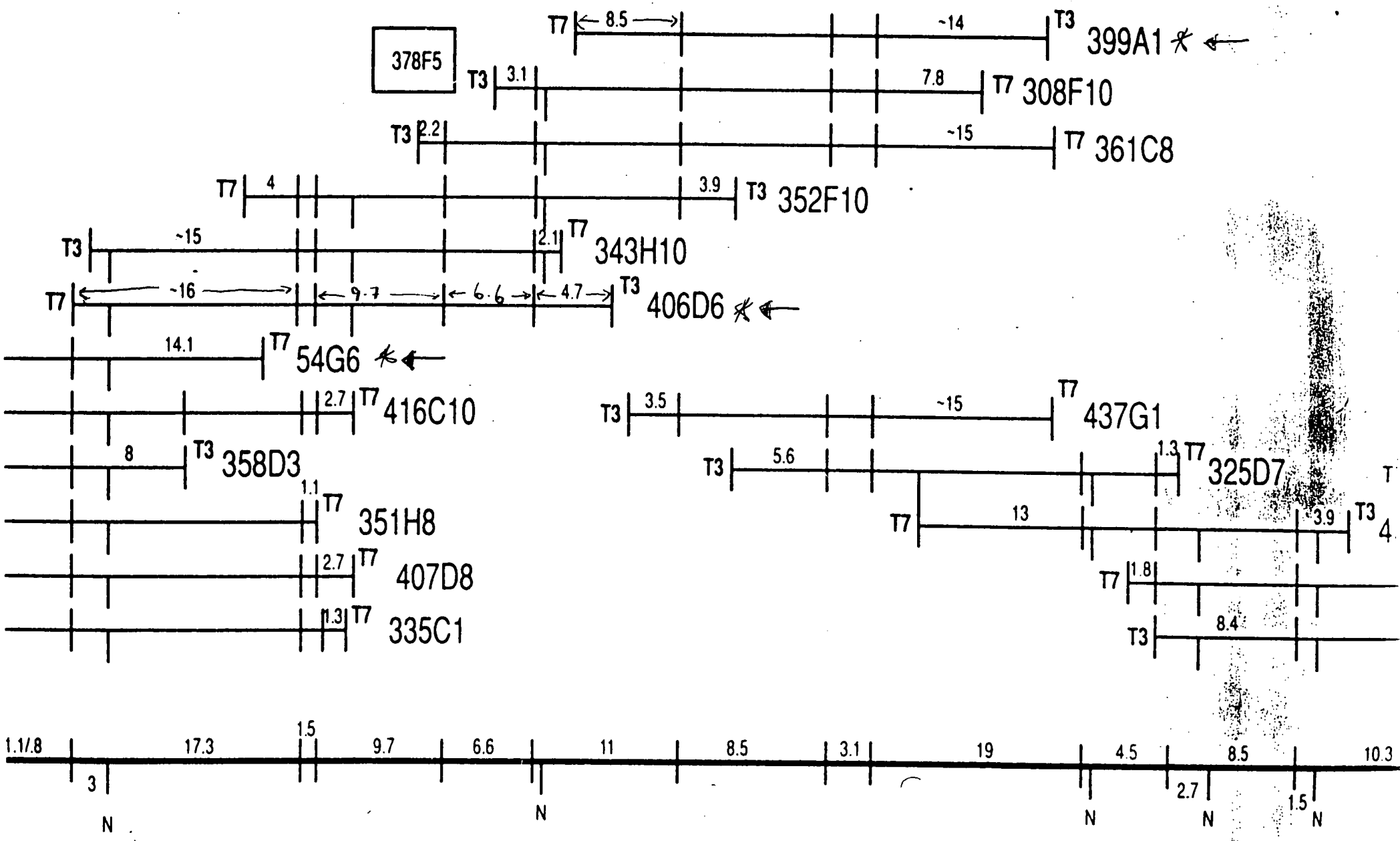

378F5
T7 8.5 ~14 T3 399A1
T3 3.1 7.8 T7 308F10
T3 2.2 ~15 T7 361C8
T7 4 3.9 T3 352F10
T3 ~15 2.1 T7 343H10
T7 ~16 9.7 6.6 4.7 T3 406D6
14.1 T7 54G6
2.7 T7 416C10
8 T3 358D3
1.1 T7 351H8
2.7 T7 407D8
1.3 T7 335C1
T3 3.5 ~15 T7 437G1
T3 5.6 1.3 T7 325D7
T7 13 3.9 T3
T7 1.8
T3 8.4
1.1/.8 17.3 1.5 9.7 6.6 11 8.5 3.1 19 4.5 8.5 10.3
3
N
N
N
N
2.7
1.5
N
6S246

FIGURE 15

FIGURE 16

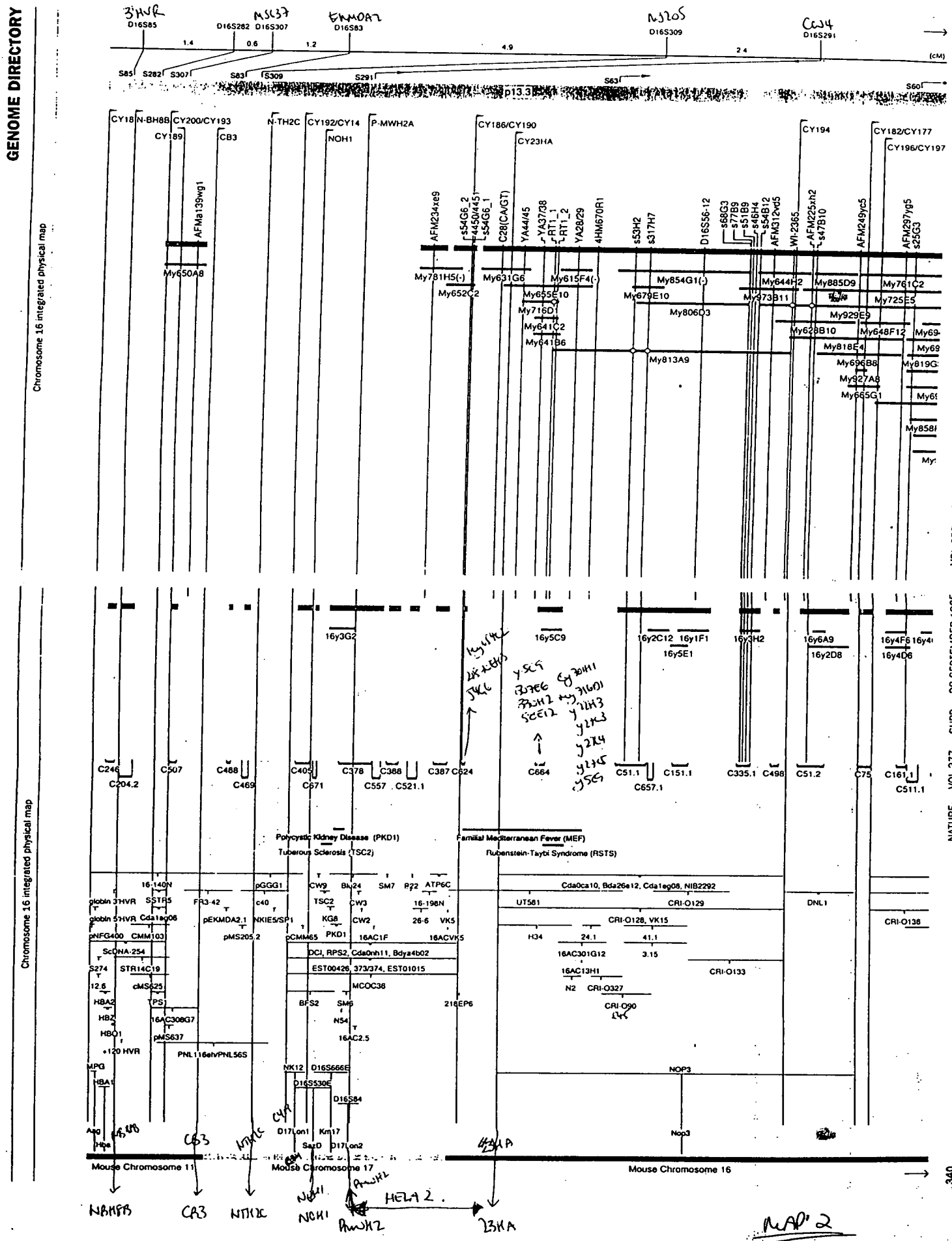
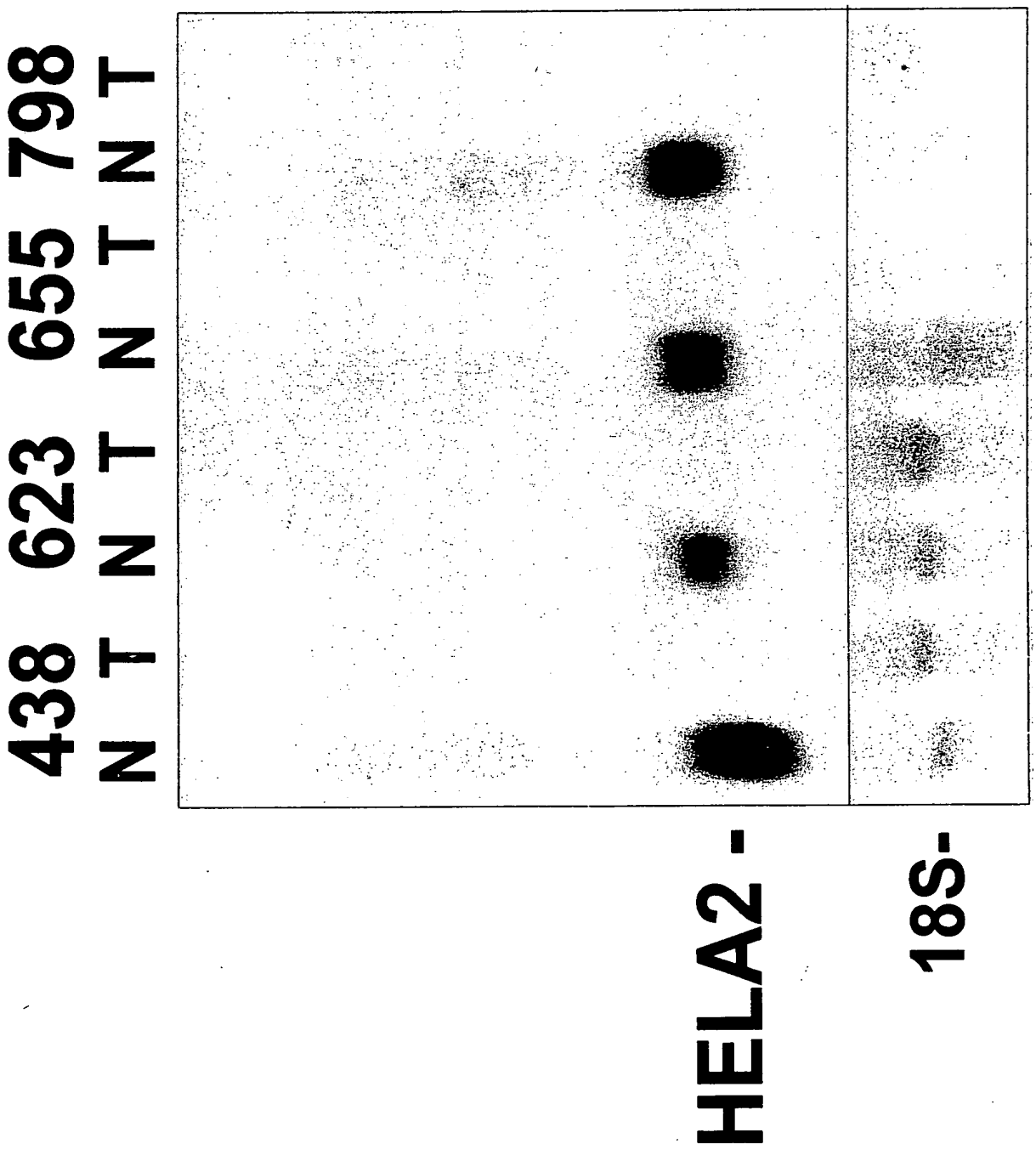
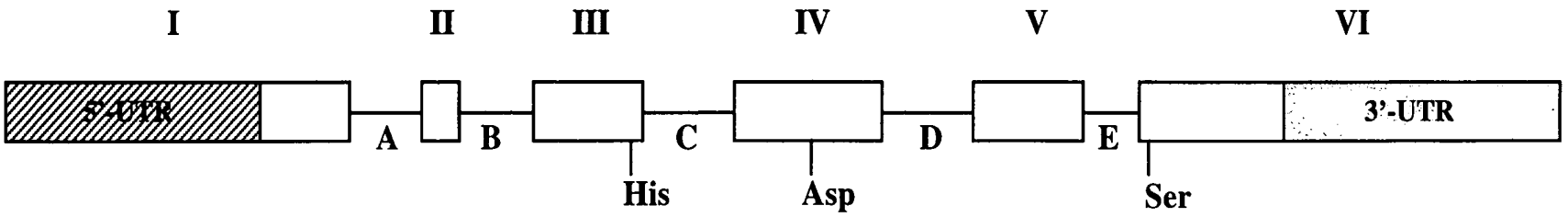

GENOME DIRECTORY
Chromosome 16 integrated physical map
3'HVR
D16S85
MSC37
D16S282
D16S307
ENMOA7
D16S83
NJ205
D16S309
CW4
D16S291
341
(CM)
S85
S282
S307
S83
S309
S291
S63
S60
p13.3
CY18
N-BH8B
CY200/CY193
CY189
CB3
N-TH2C
CY192/CY14
NOH1
P-MWH2A
CY186/CY190
CY23HA
CY194
CY182/CY177
CY196/CY197
AFM234xe9
s54G6_2
f4450/4451
s54G6_1
C28(CA/GT)
YA44/45
YA37/38
RT1_1
RT1_2
YA28/29
4HM670R1
s53H2
s317H7
D16S56-12
s68G3
s77B9
s51B9
s46H4
s54B12
AFM312vd5
WI-2365
AFM225xh2
s47B10
AFM249yc5
AFM297yg5
s25G3
AFMa139wg1
My650A8
My781H5(-)
My652C2
My631G6
My655E10
My716D1
My641C2
My641B6
My615F4(-)
My854G1(-)
My679E10
My806D3
My813A9
My644H2
My973B11
My885D9
My761C2
My725E5
My929E9
My623B10
My648F12
My818E4
My696B8
My927A8
My665G1
My819G
My858
16y3G2
16y5C9
16y2C12
16y1F1
16y5E1
16y3H2
16y6A9
16y2D8
16y4F6
16y4D6
C246
C204.2
C507
C488
C469
C405
C671
C378
C388
C557
C521.1
C387
C624
C664
C51.1
C151.1
C657.1
C335.1
C498
C51.2
C75
C161.1
C511.1
Polycystic Kidney Disease (PKD1)
Tuberous Sclerosis (TSC2)
Familial Mediterranean Fever (MEF)
Rubenstein-Taybi Syndrome (RSTS)
16-140N
globin 3'HVR
SSTR5
globin 5'HVR
Cda1eg06
pNFG400
CMM103
ScDNA-254
S274
STR14C19
12.6
cMS625
HBA2
TPS1
HBZ
16AC308G7
HBQ1
pMS637
+120 HVR
PNL116ehrPNL56S
MPG
HBA1
Apo
Hpa
pGGG1
FR3-42
c40
pEKMDA2.1
NKIE5/SP1
pMS205.2
CW9
Bw24
SM7
P22
ATP6C
TSC2
CW3
16-198N
KG8
CW2
26-6
VK5
PKD1
16AC1F
16ACVK5
pCMM65
DCI, RPS2, Cda0nh11, Bdya4b02
EST00426, 373/374, EST01015
MCOC36
BFS2
SM6
N54
16AC2.5
216EP6
NK12
D16S666E
D16S530
D16S84
D17.on1
Km17
SazD
D17.on2
Cda0ca10, Bda26e12, Cda1eg08, NIB2292
UT581
CRI-O129
DNL1
CRI-O128, VK15
CRI-O136
H34
24.1
41.1
16AC301G12
3.15
16AC13H1
CRI-O133
N2
CRI-O327
CRI-O90
NOP3
Nop3
Mouse Chromosome 11
Mouse Chromosome 17
Mouse Chromosome 16
340
NBH8B
CB3
NTHC
NOH1
PMWH2
HELA 2.
23HA
MAP 2
NATURE · VOL 377 · SUPP · 28 SEPTEMBER 1995

FIGURE 17





INTRON	TESTISIN (base pairs)	PROSTASIN (base pairs)
A	ND	243
B	ND	1763
C	716	271
D	*ca* 1800	85
E	256	92

EXON	TESTISIN (base pairs)	PROSTASIN (base pairs)
I	ND	417
II	ND	18
III	ND	163
IV	287	272
V	155	167
VI	ND	899

FIGURE 18

EXON III CACTGCTTTGAAAC**gt**gagtgggggtgcgaacggag
gggtgcggggacgggcaggaacagggctggagggagtgccaccga
actttacctctggtctgatgccagacttgggcgtgaaagttgtgc
gtggatgcggcctggtgttctcctgagccccaggctgtgctgcag
ccggttacacccactccagttccctttgggtctcctggagggaac
cctgttcaggttattccagaatgttcttccagaacatttccacac
acttttgggtattctctccctttttctttcaacccaaagttcacc
actgaccatcccaccctcatccccctcctggtggacggtgcggt
acagtgtggggcactgagccaaggccagcacccccgggccgctgt
·····INTRON C (716 BP)·····
gtggactccatcctgccaatcccacattggcgtggtgcatctccc
cattcctccttgggctgcatgggggtgcccctggaggccttggct
caatgcaaggctccttgggacagctctgggaggtgacaagacccc
acccttctgctgcaggagcaggtcctagactttggttgtggtctg
tctgggctccttcatttctgcaggggaccctgggtgttagcaagt
agcagcaacaccacagtttcccctcctgcactggaccccagttgt
gctcaggtagccagccctccatccagggcccctgactgctctctt
ctcttctgcc**ag**ctat**ag**TGACCTTAGTGATCCC EXON IV

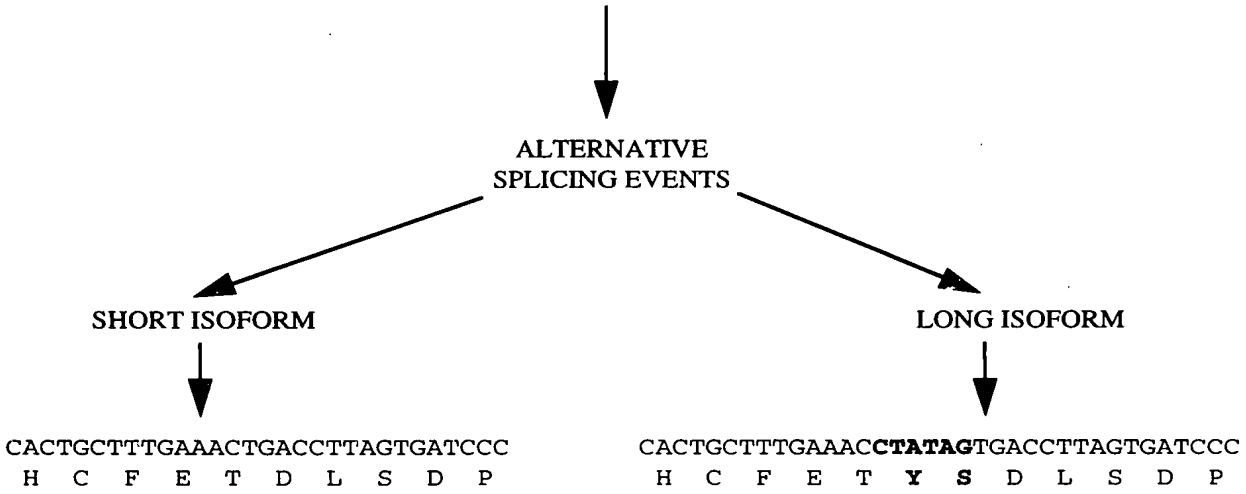


FIGURE 19

TESTISIN PRIMERS

Exonic Primers

HELA2 P1	5′ CGA AGT AAC GGG TGT AGT AG 3′	(base 379; Reverse; 20 mer; Tm 57°)
HELA2 P2	5′ TCT GTC CGG TTC TCA AAC TC 3′	(base 527; Reverse; 20 mer; Tm 61°)
HELA2 P3	5′ ACT CCG GCT TCC TGA GTC CAG C 3′	(base 85; Reverse; 22 mer; Tm 73°)
HELA2 P4	5′ CGC ATG GTC CTG ATA ACG GC 3′	(base 118; Reverse; 20 mer; Tm 70°)
HELA2 P5	5′ ACT CTA TGT GCA ACC ACC TC 3′	(base 620; Forward; 20 mer; Tm 58°)
HELA2 P6	5′ TCC TCA AGT ACA GTT TCC GC 3′	(base 640; Forward; 20 mer; Tm 62°)
HELA2 P7	5′ GGG ATC ACT AAG GTC ACT AT 3′	(base 297; Reverse; 20 mer; Tm 54°)
HELA2 P8	5′ CTG ACT TCC ATG CCA TCC TT 3′	(base 325; Forward; 20 mer; Tm 64°)
HELA2 P9	5′ GCT CAC GAC TCC AAT CTG AT 3′	(base 789; Reverse; 20 mer; Tm 61°)
HELA2 P12	5′ CAT CCT TCT GGA GCC TGC AGG 3′	(base 338; Forward; 21 mer; Tm 71°)
HELA2 P13	5′ ACA GGT GCA GAC AGC TTC ACC 3′	(base 458; Reverse; 21 mer; Tm 66°)
HELA2 P14	5′ ACG TAT GCG GAG TGA GCC TGC T 3′	(base 212; Forward; 22 mer; Tm 73°)
HELA2 P15	5′ TCA GGA AGC CGG AGT CGC AG 3′	(base 71; Forward; 20 mer; Tm 72°)
HELA2 P16	5′ CTG CGA CTC CGG CTT CCT GA 3′	(base 90; Reverse; 20 mer; Tm 72°)
HELA2 P17	5′ CTG CAA GGC ATC AAC TGG AA 3′	(base 1063; Reverse; 20 mer; Tm 66°)

NB: HELA2 **P7** includes 4 bases of **splice variant**
will work on **genomic DNA** and **long isoform cDNA**
will **NOT** work on short isoform cDNA
HELA2 **P12** and **P13** from Rachael Daniels

Intronic Primers

HELA2 P10	5′ TTA CCT CTG GTC TGA TGC CA 3′	(Forward; 20 mer; Tm 62°)
HELA2 P11	5′ GCA GAA GAG AAG AGA GCA GT 3′	(Reverse; 20 mer; Tm 57°)

FIGURE 20

```
        10          20          30          40          50          60
         *           *           *           *           *           *
CGACCTATTG TCAGGGCCCT GCGGTCACAG GACCATCCCT TCCCGTATAG TGGGTGGCGA
  D  L  L   S  G  P   C  G  H  R   T  I  P   S  R  I   V  G  G  D>

        70          80          90         100         110         120
         *           *           *           *           *           *
TGATGCTGAG CTTGGCCGCT GGGCGTGGCA AGGGAGCCTG CGTGTATGGG GCAACCACTT
  D  A  E   L  G  R   W  A  W  Q   G  S  L   R  V  W   G  N  H  L>

       130         140         150         160         170         180
         *           *           *           *           *           *
ATGTGGCGCA ACCTTGCTCA ACCGCCGCTG GGTGCTTACA GCTGCCCACT GCTTCCAAAA
  C  G  A   T  L  L   N  R  R  W   V  L  T   A  A  H   C  F  Q  K>

       190         200         210         220         230         240
         *           *           *           *           *           *
GGATAACGAT CCTTTTGACT GGACAGTCCA GTTTGGTGAG CTGACTTCCA GGCCATCTCT
  D  N  D   P  F  D   W  T  V  Q   F  G  E   L  T  S   R  P  S  L>

       250         260         270         280         290         300
         *           *           *           *           *           *
CTGGAACCTA CAGGCCTATT CCAACCGTTA CCAAATAGAA GATATTTTCC TGAGCCCCAA
  W  N  L   Q  A  Y   S  N  R  Y   Q  I  E   D  I  F   L  S  P  K>

       310         320         330         340         350         360
         *           *           *           *           *           *
GTACTCGGAG CAGTATCCCA ATGACATAGC CCTGCTGAAG CTGTCATCTC CAGTCACCTA
  Y  S  E   Q  Y  P   N  D  I  A   L  L  K   L  S  S   P  V  T  Y>

       370         380         390
         *           *           *
CAATAACTTC ATCCAGCCCA TCTGCCTCCT GAACTCCAC
  N  N  F   I  Q  P   I  C  L  L   N  S  T>
```

FIGURE 21